

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 21:22:46 ; Search time 8287 Seconds  
(without alignments)  
11386.242 Million cell updates/sec

Title: US-09-914-958B-35

Perfect score: 2177  
Sequence: 1 cggaggctcgaggccagcc.....tcctcttgaaaaaaaaa 2177

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2177	100.0	2177	6	BD269644 Human sec
2	1994	91.6	2089	9	AX357116 Sequence
3	1891	86.9	2681	9	AJ276171 Homo sapi
4	1759	80.8	2589	6	AX357100 Sequence
5	1758	80.8	2039	6	AX357115 Sequence
6	1758	80.8	2145	6	AX357117 Sequence
7	1745	80.2	1959	6	AX470036 Sequence
8	1745	80.2	2507	6	AX470035 Sequence
9	1745	80.2	2507	9	AJ279016 Homo sapi
10	1744	80.1	1947	6	AX470056 Sequence
11	1744	80.1	1947	9	AJ21515 Homo sapi
12	1687	77.5	2889	9	AK057190 Homo sapi
13	1617	74.3	2413	6	AX427468 Sequence
14	1438	66.1	2147	9	BC034245 Homo sapi
15	1428	65.6	2263	6	AX202100 Sequence
16	1314	60.4	1501	6	AR339541 Sequence
17	1236	56.8	2178	6	AX876177 Sequence
18	1236	56.8	2178	6	BD155992 Primer fo
19	1236	56.8	2178	9	AK001182 Homo sapi
20	738	33.9	789	6	AR379939 Sequence
21	648	29.8	1143	6	BD205640 97 human
22	496	22.8	608	6	AX867169 Sequence
23	496	22.8	608	6	BD147231 Primer fo
24	225	10.3	146124	9	AL358938 Human DNA
25	203	9.3	161235	9	AL139239 Human DNA
26	203	9.3	172383	2	AC015630 Homo sapi
27	160	7.3	172383	2	AC015630 Homo sapi
28	143	6.6	409	11	HSJ123C5
29	113	5.2	395	11	G25373
30	113	5.2	492	11	G27517
31	109	5.0	707	6	AR379664 Sequence
32	57	2.6	57	6	AX470037
33	54	2.5	1558	10	BC024472 Mus muscu
34	54	2.5	1973	10	MMU421516 Mus muscu
35	54	2.5	163348	10	AC119236 Mus muscu
36	54	2.5	189196	10	AL603804 Mouse DNA
37	44	2.0	700	10	RN078304 Rattus norv
38	40	1.8	752	6	AX202149 Sequence
39	30	1.4	210385	10	AC106128 Rattus no
40	30	1.4	243770	2	AC120763 Rattus no
41	30	1.4	252404	2	AC097117 Rattus no
42	29	1.3	164766	10	AC144792 Mus muscu
43	25	1.1	25	6	AX357106 Sequence
44	24	1.1	2012	8	AY237400 Malus x d
45	23	1.1	23	6	AX357108 Sequence

ALIGNMENTS

RESULT 1	BD269644	2177 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD269644				
DEFINITION	Human secretory proteins.				
ACCESSION	BD269644				
VERSION	BD269644.1	GI:33079412			
KEYWORDS	JP 2002537805-A/13.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 2177)				
AUTHORS	Tang, T. Y., Lal, P., Baughn, M. R., Yue, H., Young, J. A., Lu, D. A. M. and				
	Azimzai, Y.				
TITLE	Human secretory proteins				





Db	1681	GGGAGATGAAC	T	CAGTGTG	GAGATCCT	CTA	CCCCCGGGATGAGGACAC	CACTTCAGGACC	1744
Qy	1741	CAGCCCACTGGAG	T	GTGGCCAA	GAGATTCT	CC	CAGCAGGAAAAATGGCCAT	TTGCATGGACA	1800
Db	1741	CAGCCCACTGGAG	T	GTGGCCAA	GAGATTCT	CC	CAGCAGGAAAAATGGCCAT	TTGCATGGACA	1800
Qy	1801	CCAATGAATGCAT	C	CAAGTTCC	CAATTT	CGTGTG	CCCTCGAGACAAGCCCGTATGTGTCAACA	1860	
Db	1801	CCAATGAATGCAT	C	CAAGTTCC	CAATTT	CGTGTG	CCCTCGAGACAAGCCCGTATGTGTCAACA	1860	
Qy	1861	CCTATGGAGCTAC	AGTGC	CGGACCA	CAAGAA	GTGTC	AGTGCAGTCCGGGCTACGAGCCCAACG	1920	
Db	1861	CCTATGGAGCTAC	AGTGC	CGGACCA	CAAGAA	GTGTC	AGTGCAGTCCGGGCTACGAGCCCAACG	1920	
Qy	1921	AGGATGSCACAG	C	CGTGGCTG	GTGGAG	CCCTGT	GTGTTGAAGATAGTACACCAACG	1980	
Db	1921	AGGATGSCACAG	C	CGTGGCTG	GTGGAG	CCCTGT	GTGTTGAAGATAGTACACCAACG	1980	
Qy	1981	TTGGGAAGAGC	CTTGGT	CCCTGAA	TCACTG	AAATCA	CTGAAATCAACCGCTTGAATA	2040	
Db	1981	TTGGGAAGAGC	CTTGGT	CCCTGAA	TCACTG	AAATCA	CTGAAATCAACCGCTTGAATA	2040	
Qy	2041	CTGTGTCATCAG	AAACA	CTTACCTG	GGAACCTT	CACTG	AGCAGGATACAAAATTTCCATTTC	2100	
Db	2041	CTGTGTCATCAG	AAACA	CTTACCTG	GGAACCTT	CACTG	AGCAGGATACAAAATTTCCATTTC	2100	
Qy	2101	TTAAGCTATTAA	TACATTA	AGATTG	GGGGTG	TACCTT	ACATAATAAAATTTCCATTTC	2160	
Db	2101	TTAAGCTATTAA	TACATTA	AGATTG	GGGGTG	TACCTT	ACATAATAAAATTTCCATTTC	2160	
Qy	2161	TCTTGA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	2177	
Db	2161	TCTTGA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	2177	
RESULT 2									
AX357116									
LOCUS	AX357116								
DEFINITION	Sequence 17 from Patent WO0206478.								
ACCESSION	AX357116								
VERSION	AX357116.1								
KEYWORDS	GI:18674296								
SOURCE	Homo sapiens (human)								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.								
AUTHORS	Heus, H.C., Nelissen, R.L. and Meeuwisse, C.M.								
TITLE	Extracellular matrix protein								
JOURNAL	Patent: WO 0206478-A 17 24-JAN-2002;								
FEATURES	Akzo Nobel N.V. (NL)								
source	Location/Qualifiers								
	1..2089								
	/organism="Homo sapiens"								
	/mol_type="unassigned DNA"								
	/db_xref="taxon:9606"								
ORIGIN									
Query Match	91.6%;								
Best Local Similarity	100.0%;								
Matches 2044;	Conservative								
	0; Mismatches								
	1; Indels								
	0; Gaps								
Qy	133	GGCTGAC	CCCGGCA	TGTC	CCAGGATG	TACCGT	TCCCTGCTGCTGCTCTGGTTTCTGCCCA	192	
Db	45	GGCTGAC	CCCGGCA	TGTC	CCAGGATG	TACCGT	TCCCTGCTGCTGCTCTGGTTTCTGCCCA	104	
Qy	193	TCAC	TGAGGGT	CC	CAGGGG	GTGA	ACCCATGTTCACTGCACTCAACCACTCAGTTCTGC	252	
Db	105	TCAC	TGAGGGT	CC	CAGGGG	GTGA	ACCCATGTTCACTGCACTCAACCACTCAGTTCTGC	164	
Qy	253	CTC	TGACTAT	GACAG	TAAAT	CC	CACCCAGCTCAACTATGGTGGCAGTTACTGATGG	312	
Db	165	CTC	TGACTAT	GACAG	TAAAT	CC	CACCCAGCTCAACTATGGTGGCAGTTACTGATGG	224	

QY	1393	ATGAGAGTCCATGGCTCAGCGCTGTCCTTCCTCGGGCAATCAGGGCTTCAACAACA	1452
Db	1305	ATGAGAGTCCATGGCTCAGCGCTGTCCTTCCTCGGGCAATCAGGGCTTCAACAACA	1364
QY	1453	ACTGGTGCAGTGGTGCACGACCCCGTTTGGGGCTTTTGGCGGAGCTAAAGTGC	1512
Db	1365	ACTGGTGCAGTGGTGCACGACCCCGTTTGGGGCTTTTGGCGGAGCTAAAGTGC	1424
QY	1513	TGCTCTACACCAAGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTAAC	1572
Db	1425	TGCTCTACACCAAGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTAAC	1484
QY	1573	TGTGTGAGATGAGAGCCGTCGACACATTTGGCTTGGGAAAGGATGAAGCCAGAGTGG	1632
Db	1485	TGTGTGAGATGAGAGCCGTCGACACATTTGGCTTGGGAAAGGATGAAGCCAGAGTGG	1544
QY	1633	AGTGACGTGGCCAGATGGCAAGATGGTGAAGCCGGAACGTGGCCAGCGGGAGATGAAC	1692
Db	1545	AGTGACGTGGCCAGATGGCAAGATGGTGAAGCCGGAACGTGGCCAGCGGGAGATGAAC	1604
QY	1693	CAGTGTGGAGATCCCTTACCCCGGATGAGACACATTCAGGACCCAGCCCACTGG	1752
Db	1605	CAGTGTGGAGATCCCTTACCCCGGATGAGACACATTCAGGACCCAGCCCACTGG	1664
QY	1753	AGTGTGCCAAGATTTCTCCAGCAGGAAATGGCCATTCATGGACACCAATCAATGCA	1812
Db	1665	AGTGTGCCAAGATTTCTCCAGCAGGAAATGGCCATTCATGGACACCAATCAATGCA	1724
QY	1813	TCAGTTCCTTCCATTCGTCGCTTCGAGACAAAGCCCGTATGTGTCAACACCTATGGAAGCT	1872
Db	1725	TCAGTTCCTTCCATTCGTCGCTTCGAGACAAAGCCCGTATGTGTCAACACCTATGGAAGCT	1784
QY	1873	ACAGGTGCCGACCAACAGAGTGCAGTCGGGGCTACGACCCAGAGGATGGACAG	1932
Db	1785	ACAGGTGCCGACCAACAGAGTGCAGTCGGGGCTACGACCCAGAGGATGGACAG	1844
QY	1933	CCTGCTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCAAGTTGGGAAGCC	1992
Db	1845	CCTGCTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCAAGTTGGGAAGCC	1904
QY	1993	TTGTTCCCTGAATCACTGAATCACTGCTTGAATCAACCGCTGGAATACCTGTTGATCAG	2052
Db	1905	TTGTTCCCTGAATCACTGAATCACTGCTTGAATCAACCGCTGGAATACCTGTTGATCAG	1964
QY	2053	GAACTTACCTGGACCTTCACTGACGAGATCAACCTTCTATTGATTAAGCTATTAA	2112
Db	1965	GAACTTACCTGGACCTTCACTGACGAGATCAACCTTCTATTGATTAAGCTATTAA	2024
QY	2113	TACATTAAGATTGGGGTGTACCTTACATTAATAATTCCTTCTTGAAAAAA	2172
Db	2025	TACATTAAGATTGGGGTGTACCTTACATTAATAATTCCTTCTTGAAAAAA	2084
QY	2173	AAAAA 2177	
Db	2085	AAAAA 2089	

RESULT 3  
HSA276171  
LOCUS HSA276171 2681 bp mRNA linear PRI 20-JUL-2000  
DEFINITION Homo sapiens mRNA for ASPIC (acidic secreted protein in cartilage) (ASPIC1 gene).  
ACCESSION AJ276171  
VERSION AJ276171.1 GI:9368805  
KEYWORDS acidic secreted protein in cartilage; ASPIC1 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
Bolton, M.C., Wait, R. and Saklatvala, J.  
AUTHORS Cloning of ASPIC, a novel protein secreted by human normal and  
TITLE osteoarthritic cartilage, identified by 2D electrophoresis and mass

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 2681)  
Bolton, M.C.  
Direct Submission

Submitted (17-JUL-2000) Bolton M.C., Cell signalling, Kennedy  
Institute of Rheumatology, 1 Aspenlea Road, Hammersmith, London W6  
8UH, UNITED KINGDOM

FEATURES  
source

1..2681  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_type="chondrocyte"  
/tissue\_type="cartilage"  
1..2681  
/gene="ASPIC1"  
1..149  
/gene="ASPIC1"  
/note="A second transcription start site at +31"  
/evidence=experimental  
150..2135  
/gene="ASPIC1"  
/standard\_name="acidic secreted protein in cartilage"  
/codon\_start=1  
/evidence=experimental  
/product="ASPIC"  
/protein\_id="CAB98267.1"  
/db\_xref="GI:9368806"  
/db\_xref="GOA:Q9NQ80"  
/db\_xref="SPTREMBL:Q9NQ80"  
/translation="NAPSADPGSRMLPFLILLWFLPITEGSRQAEPMFTAVTNSVLP  
PYDLSNPTQINLYGAVTDVDDHDFEIVVAGYNGPNLVKYDRAQRLVNIAYDERS  
PYALDRQNAIGVTACDIDGDEEYFLNTNNAFSGVATYDKLKFPRNRWEDI  
LSEVNVARGVASLFAGRSACVDRKSGRYSTIYANYAGNVGPDALIMDEPADSL  
SGILALRVAAGAGSVKTYGGRVSGVPISSASDI FCDNENGPNFLHNGDGTFF  
VDAASAGVDDPHQGRGVALADFNEDGKVDIVYGNWGNPHRLYLQMSHGKVPFRI  
ASPKSPSPSVRTVITADFDNDQELIEIFNNIAYSSANRLFRVIRRHGDPLIEL  
RTRGAFARGAKVLYTKSGAHLRIIDGSGYLCEMEPVAFHGLGKEASSVEVTWP  
DGKVMYRNVASGEMNSVLEILYPRDEDTLQDPAPLECGQGFQOENHGMCDTNECLOF  
PVCPRDKPVCVNTYGRCTNKKCSRGVEPNEDGTACVGLGQSGPGRPTTPTTAA  
ATAAAAAAGATAAPLVLDGDLNLSGVKESCEPSC"

gene

5'UTR

CDS

sig\_peptide

CDS

150..230  
/gene="ASPIC1"  
174..2135  
/gene="ASPIC1"  
/standard\_name="acidic secreted protein in cartilage"  
/note="putative alternative start, less identity to the  
Kozak rule"  
/codon\_start=1  
/evidence=experimental  
/product="ASPIC"  
/protein\_id="CAB98268.1"  
/db\_xref="GI:9368807"  
/db\_xref="GOA:Q9NQ79"  
/db\_xref="SPTREMBL:Q9NQ79"

174..230  
/gene="ASPIC1"  
/note="putative alternative start, less identity to the  
Kozak rule"

sig\_peptide

174..230  
/gene="ASPIC1"  
/note="putative alternative start, less identity to the  
Kozak rule"

```

CDS
183..2135
/genes="ASPI1"
/standard name="acidic secreted protein in cartilage"
/notes="putative alternative start, less identity to the Kozak rule"
/codon_start=1
/evidence=experimental
/product="ASPI1"
/protein_id="CAB98269.1"
/db_xref="GI:936808"
/db_xref="GOA:Q9NQ78"
/db_xref="SPREBEL:Q9NQ78"
/translation="MUPFLLLLWFLPITGSGQRAEMFTAVTNSVLPPDYSNPQLN
YGVAVTDVDDGDEFIVAGYNGPNLVLYDRAQKLVNIADVDSPIYALDRQGN
AIGVTACDIDGGERIEYFLINTNARFSGVATYDKLFKFRNNRWEDILSDVNVARGV
ASLFAGRSVACVDRKSGRYSIVIANAYGVNVPDALIEMDPEASLGRILALROVA
ARAGVSKYTGGRGVSGPILSSASDIFCDNENGNPFLPHNRGDGTFVDAASAGVDD
PIQHGRGVALADPNRQKVDIVYGNWNGPHRLYLQMSHGKVRFDIAIPKFSMPSPV
RVTTFADFNQDELEIFFNIAVRSANRKFVIRREHGDPLIEELNPDALPEGR
GTGVVTFDGDGMLDLILSHGESMAQPLSVFRGNQFNMMRLVVRTRFGAFARGA
KVLVYTKSGAHLRIIDGSGYLCEMEPVAFHGLGKDEASVVEVTFDGMKVMRNVAS
GEMNSVLEILYPRDEDTLODPAPLECGQGFSGOENGHCHMDNEICIOFPFVCPDRKPV
VNTYGSVCRCTNKKCSRGVEPNEDGTACVGTILQSGSPRPTTTTAAATAAAAAAGA
ATRAAPVLVGDNLGSLVVKESCEPSC"
183..230
sig_peptide
/genes="ASPI1"
/notes="putative alternative start, less identity to the Kozak rule"
231..2132
mat_peptide
/genes="ASPI1"
/product="ASPI1"
/standard name="acidic secreted protein in cartilage"
2136..2681
3'UTR
/genes="ASPI1"
2642..2647
polyA_signal
/genes="ASPI1"

ORIGIN
Query Match 86.9%; Score 1891; DB 9; Length 2681;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGAGGCTCGAGGCGAGCGCGCGCTCCACGCCCTTAGCGGCGGGCGAGAGCGG 60
DB 28 CGGAGGCTCGAGGCGAGCGCGCGCTCCACGCCCTTAGCGGCGGGCGAGAGCG 87
QY 61 GCAGAGCGGCGAGGCGCGCGCTCCACGCCCTTAGCGGCGGGCGAGAGCGGGA 120
DB 88 GCAGAGCGGCGAGGCGCGCGCTCCACGCCCTTAGCGGCGGGCGAGAGCGGGA 147
QY 121 GGATGCTCGAGCGCTGACCCCGCATGTCCAGGATGTTACCGTTCTCTGCTCT 180
DB 148 GGATGCTCGAGCGCTGACCCCGCATGTCCAGGATGTTACCGTTCTCTGCTCT 207
QY 181 GFTTTCGCCATCACTAGGAGGTCCAGCGGCTGAACCCATGTTCACTGCGATCA 240
DB 208 GFTTTCGCCATCACTAGGAGGTCCAGCGGCTGAACCCATGTTCACTGCGATCA 267
QY 241 ACTCAGTTCGCTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTG 300
DB 268 ACTCAGTTCGCTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTG 327
QY 301 TTAAGTATGAGGACATGATGGGACATTTGAGATCGTGTGGGGGTACAATGGACC 360
DB 328 TTAAGTATGAGGACATGATGGGACATTTGAGATCGTGTGGGGGTACAATGGACC 387
QY 361 ACCTGTTCTGAGATATGACGGGCCAGAGGGCTGTTGATCATCGCGGTGATGAG 420
DB 388 ACCTGTTCTGAGATATGACGGGCCAGAGGGCTGTTGATCATCGCGGTGATGAG 447
QY 421 GCAGCTCACCTACTACGCGCTCGGACCGGCGAGGGGACGCCATTTGGGGTCA 480
DB 448 GCAGCTCACCTACTACGCGCTCGGACCGGCGAGGGGACGCCATTTGGGGTCA 507

```

```

QY 481 GCGACATCGACGGGAGCGCGGAGGAGATCTACTTCTCAACACCAATAATGCTCTT 540
DB 508 GCGACATCGACGGGAGCGCGGAGGAGATCTACTTCTCAACACCAATAATGCTCTT 567
QY 541 CGGGGGTGGCCAGGTACACCGACAGTGTCTCAAGTTCGGCAATAACGGTGGGAG 600
DB 568 CGGGGGTGGCCAGGTACACCGACAGTGTCTCAAGTTCGGCAATAACGGTGGGAG 627
QY 601 TCTGAGCGCATGAGGTCAACGTCGCCCGTGTGTGCGGAGCCTCTTTGCCGAGC 660
DB 628 TCTGAGCGCATGAGGTCAACGTCGCCCGTGTGTGCGGAGCCTCTTTGCCGAGC 687
QY 661 TGGCCTGTGTGGACAGAAAGGGCTGTGGACGCTACTCTATPCTACATTCGCAAT 720
DB 688 TGGCCTGTGTGGACAGAAAGGGCTGTGGACGCTACTCTATPCTACATTCGCAAT 747
QY 721 ACGGTAATGTGGCCCTGATGCCCTCAATTGAATGACACCTGAGGCGAGTCTCTCC 780
DB 748 ACGGTAATGTGGCCCTGATGCCCTCAATTGAATGACACCTGAGGCGAGTCTCTCC 807
QY 781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAAGCAATATAC 840
DB 808 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAAGCAATATAC 867
QY 841 GCGAGCGCTAGCGTGGGCGCCATCTCAGCAGAGTGTGCTCGGATATCTTCTGAG 900
DB 868 GCGAGCGCTAGCGTGGGCGCCATCTCAGCAGAGTGTGCTCGGATATCTTCTGAG 927
QY 901 ATGAGATGGGCTTAATCTTCTTTCACACCGGGCGAGTGTGCTGAGCGCTG 960
DB 928 ATGAGATGGGCTTAATCTTCTTTCACACCGGGCGAGTGTGCTGAGCGCTG 987
QY 961 CGGCCAGTGTGTGTGGACGACGCCCATCCAGCATGGGCGAGTGTGCTGAGCT 1020
DB 988 CGGCCAGTGTGTGTGGACGACGCCCATCCAGCATGGGCGAGTGTGCTGAGCT 1047
QY 1021 TCAACCGTGTGTCGAGTGGACATCGTGTATGGCAATGGGCGGCGGCGCTCT 1080
DB 1048 TCAACCGTGTGTCGAGTGGACATCGTGTATGGCAATGGGCGGCGGCGCTCT 1107
QY 1081 ATCTGCAATGACACCGCATGGAGTGTGCTTCCGGGACATGCTTCAACCAAGTCT 1140
DB 1108 ATCTGCAATGACACCGCATGGAGTGTGCTTCCGGGACATGCTTCAACCAAGTCT 1167
QY 1141 CCATGCGCTCTCTGTCGCGACCGTCAATCCCGCGATTTTGACATGACAGAGCTG 1200
DB 1168 CCATGCGCTCTCTGTCGCGACCGTCAATCCCGCGATTTTGACATGACAGAGCTG 1227
QY 1201 AGATCTTCTTCAACCAATTTGCTACCGAGTCTCTCAGCGCAACCGCTCTTCCG 1260
DB 1228 AGATCTTCTTCAACCAATTTGCTACCGAGTCTCTCAGCGCAACCGCTCTTCCG 1287
QY 1261 TCCGTAGAGACGAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1320
DB 1288 TCCGTAGAGACGAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1347
QY 1321 CTGAGGCGGCGGCGACAGGGGGTGTGTGACCACTTCGACGGAGACGGGATGTGAG 1380
DB 1348 CTGAGGCGGCGGCGACAGGGGGTGTGTGACCACTTCGACGGAGACGGGATGTGAG 1407
QY 1381 TCATCTTGTCCCATGAGAGTTCATGGCTGAGCGGCTGTGCTTCCGGGGCAATCAG 1440
DB 1408 TCATCTTGTCCCATGAGAGTTCATGGCTGAGCGGCTGTGCTTCCGGGGCAATCAG 1467
QY 1441 GCTTCAACCAACTGCTGCGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
DB 1468 GCTTCAACCAACTGCTGCGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCG 1527
QY 1501 GAGCTAAGTGTGTCTTACACCAAGAGTGGGCGGCGGCGGCGGCGGCGGCGG 1560
DB 1528 GAGCTAAGTGTGTCTTACACCAAGAGTGGGCGGCGGCGGCGGCGGCGGCGG 1587

```









```
Db 1185 ACGAGACCCCTCATCGAGGAGCTCAATCCCGCGGACGCGCTTGGAGCCTGAGGGCCGGG 1244
QY 1333 GCACAGGGGTGTGGTACCGACTTCGACGAGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
Db 1245 GCACAGGGGTGTGGTACCGACTTCGACGAGAGACGGGATGCTGGACCTCATCTTGTCCC 1304
QY 1393 ATGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCGCGGGGAAATCAGGGGCTTCAACAACA 1452
Db 1305 ATGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCGCGGGGAAATCAGGGGCTTCAACAACA 1364
QY 1453 ACTGGCTGCCAGTGGTCCACGACCCGCTTTGGGGCTTTGGCCAGGGGAGCTAAGGTGG 1512
Db 1365 ACTGGCTGCCAGTGGTCCACGACCCGCTTTGGGGCTTTGGCCAGGGGAGCTAAGGTGG 1424
QY 1513 TGCTCTACCAAGAGAGTGGGCCCCACCTGAGGATCATCGACGGGGCTCAGGCTACC 1572
Db 1425 TGCTCTACCAAGAGAGTGGGCCCCACCTGAGGATCATCGACGGGGCTCAGGCTACC 1484
QY 1573 TGTGTGAGATGGAGCCGCTGGACACTTTGGCCCTGGGGAAGATGAAGCCAGCGATGG 1632
Db 1485 TGTGTGAGATGGAGCCGCTGGACACTTTGGCCCTGGGGAAGATGAAGCCAGCGATGG 1544
QY 1633 AGGTGAGTGGCCAGATGGCAAGATGTGAGCGGAACGTGGCCAGCGGGGAGATGAAT 1692
Db 1545 AGGTGAGTGGCCAGATGGCAAGATGTGAGCGGAACGTGGCCAGCGGGGAGATGAAT 1604
QY 1693 CAGTGTGAGATCCTTACCCCGGGATGAGACACACTTTCAGGACCCAGCCCCACTGG 1752
Db 1605 CAGTGTGAGATCCTTACCCCGGGATGAGACACACTTTCAGGACCCAGCCCCACTGG 1664
QY 1753 AGTGTGCCAAGATTCCTCCACAGAGAAATGGCCATTCATGGACACCAATGAATGCA 1812
Db 1665 AGTGTGCCAAGATTCCTCCACAGAGAAATGGCCATTCATGGACACCAATGAATGCA 1724
QY 1813 TCCAGTTCCTCATTCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACTATGGAAGCT 1872
Db 1725 TCCAGTTCCTCATTCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACTATGGAAGCT 1784
QY 1873 ACAGGTGCCGGAACAACAAGATGCAATCGGGCTACGAGCCCAACAGAGATGGACAG 1932
Db 1785 ACAGGTGCCGGAACAACAAGATGCAATCGGGCTACGAGCCCAACAGAGATGGACAG 1844
QY 1933 CCTGCGTGG 1941
Db 1845 CCTGCGTGG 1853

RESULT 7
AX470036
LOCUS AX470036 1959 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 2 from Patent WO02053709.
ACCESSION AX470036
VERSION AX470036.1 GI:22205283
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Richter, W. and Steck, B.
AUTHORS Cartilage marker gene, method for the production and use thereof
TITLE Patent: WO 02053709-A 2 11-JUL-2002;
JOURNAL CITONET GMBH & CO KG (DE)
FEATURES
Location/Qualifiers
source 1..1959
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 80.2%; Score 1745; DB 6; Length 1959;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 147 ATGTCCAGGATGTTACCGTTCCTGCTGCTCTGTGTTCTGCCCCATCACTGAGGGTCC 206
Db 1 ATGTCCAGGATGTTACCGTTCCTGCTGCTCTGTGTTCTGCCCCATCACTGAGGGTCC 60
QY 207 CAGGGGCTGAACCCATGTTCACTGAGTCAACAACTCAGTTCCTGCTCCTGATGAC 266
Db 61 CAGGGGCTGAACCCATGTTCACTGAGTCAACAACTCAGTTCCTGCTCCTGATGAC 120
QY 267 AGTAATCCCAACCCAGCTCAACTATGTTGGAGTCTACTGATGGAACCATGATGGGAC 326
Db 121 AGTAATCCCAACCCAGCTCAACTATGTTGGAGTCTACTGATGGAACCATGATGGGAC 180
QY 327 TTTGAGATCGTCTGGCGGGTAAATGGAACCAACCTGTTCTTGAAGTATGACCGGGCC 386
Db 181 TTTGAGATCGTCTGGCGGGTAAATGGAACCAACCTGTTCTTGAAGTATGACCGGGCC 240
QY 387 CAGAAGCGGCTGTTGAACATCGCGGTGATGAGGCGAGCTCACTACTACGCGTGGG 446
Db 241 CAGAAGCGGCTGTTGAACATCGCGGTGATGAGGCGAGCTCACTACTACGCGTGGG 300
QY 447 GACCGGAGGGGAAACGCAATTGGGGTCAACAGCTCGACATCGACGGGACGCGCGGAG 506
Db 301 GACCGGAGGGGAAACGCAATTGGGGTCAACAGCTCGACATCGACGGGACGCGCGGAG 360
QY 507 GAGATCTACTTCTCTCAACACCAATAATGCTTCTCGGGGGTGGCCAGCTACACCGACAAG 566
Db 361 GAGATCTACTTCTCTCAACACCAATAATGCTTCTCGGGGGTGGCCAGCTACACCGACAAG 420
QY 567 TTGTTCAAGTTCGGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAAGTGGCC 626
Db 421 TTGTTCAAGTTCGGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAAGTGGCC 480
QY 627 CGTGTGTGGCCAGCCTCTTTCGGGAGCTCTGTGGCCCTGTGTGGACAGAAAGGCTCT 686
Db 481 CGTGTGTGGCCAGCCTCTTTCGGGAGCTCTGTGGCCCTGTGTGGACAGAAAGGCTCT 540
QY 687 GGACGCTACTCTATCTACATTCGCAATTAAGCCCTACGTAATGTGGGCCCTGTATGCCCTC 746
Db 541 GGACGCTACTCTATCTACATTCGCAATTAAGCCCTACGTAATGTGGGCCCTGTATGCCCTC 600
QY 747 ATTGAATGGACCTCAGGCGAGTCACTTCCGGGGCAATTCCTGGGCTCAGAGATGTG 806
Db 601 ATTGAATGGACCTCAGGCGAGTCACTTCCGGGGCAATTCCTGGGCTCAGAGATGTG 660
QY 807 GCTGCTGAGGCTGGGTTCAGCAATATACAGGGGGCGAGGCGTTCAGCGTGGGCCCATC 866
Db 661 GCTGCTGAGGCTGGGTTCAGCAATATACAGGGGGCGAGGCGTTCAGCGTGGGCCCATC 720
QY 867 CTGAGCAGAGTGCCTCGGATATCTTCTGCAAAATGAGAATGGGCTAACTTCTTTTC 926
Db 721 CTGAGCAGAGTGCCTCGGATATCTTCTGCAAAATGAGAATGGGCTAACTTCTTTTC 780
QY 927 CACAACGGGGGAGTGGCACCTTTGTGGAGCTGCGGCCAGTCTGGTGTGGAGACCCC 986
Db 781 CACAACGGGGGAGTGGCACCTTTGTGGAGCTGCGGCCAGTCTGGTGTGGAGACCCC 840
QY 987 CACGACATGGGGAGGTGTGGCCCTGCTGCTCAACCGTGTGCAAGTGGACATC 1046
Db 841 CACGACATGGGGAGGTGTGGCCCTGCTGCTCAACCGTGTGCAAGTGGACATC 900
QY 1047 GTCTATGGCACTGGAATGGCCCCACCGCCCTTATCTGCAAAATGAGACCCCATGGGAAG 1106
Db 901 GTCTATGGCACTGGAATGGCCCCACCGCCCTTATCTGCAAAATGAGACCCCATGGGAAG 960
QY 1107 GTCCGCTTCGGGACATCGCCTCACCAAGTCTCCATGCCCTCCCTGTCCGACGGTC 1166
Db 961 GTCCGCTTCGGGACATCGCCTCACCAAGTCTCCATGCCCTCCCTGTCCGACGGTC 1020
QY 1167 ATCACCGCCGACTTTGACAAATGACAGGAGCTGGAGATCTTCTTCAACAACTTGCTAC 1226
Db 1021 ATCACCGCCGACTTTGACAAATGACAGGAGCTGGAGATCTTCTTCAACAACTTGCTAC 1080
```





Db	1021	ATCACCGCGAGCTTTGACAATGACACGAGAGCTGGAGATCTTCTTCAACAAATTTGCTAC	1081
QY	1227	CGAGCTCTCTCAGCAACCGGCTCTTCGGGCTCATCCGTAGAGACACGAGACCCGCTC	1386
Db	1081	CGAGCTCTCTCAGCAACCGGCTCTTCGGGCTCATCCGTAGAGACACGAGACCCGCTC	1140
QY	1287	ATCGAGAGCTCAATCCCGGGACGCTTTGAGAGCTCAGGCGCCGGGCGACAGGGGGTGTG	1346
Db	1141	ATCGAGAGCTCAATCCCGGGACGCTTTGAGAGCTCAGGCGCCGGGCGACAGGGGGTGTG	1200
QY	1347	GTGACCGACTTCGACGGAGACGGATCTGGACCTCATCTGTGCCATGGAGATCCATG	1406
Db	1201	GTGACCGACTTCGACGGAGACGGATCTGGACCTCATCTGTGCCATGGAGATCCATG	1260
QY	1407	GCTCAGCGCTGTCCGCTTTCGGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTG	1466
Db	1261	GCTCAGCGCTGTCCGCTTTCGGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTG	1320
QY	1467	GTGCCACGACCCGGTTTGGGGCTTTTGCCAGGGGAGCTAAGGTCGTCTTACACCAAG	1526
Db	1321	GTGCCACGACCCGGTTTGGGGCTTTTGCCAGGGGAGCTAAGGTCGTCTTACACCAAG	1380
QY	1527	AAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGCTACTCTGTGTGATGGAG	1586
Db	1381	AAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGCTACTCTGTGTGATGGAG	1440
QY	1587	CCGCTGGCACACATTTGGCCTTGGGGAAGGATGAACCCAGCAGTGTGGAGTGAAGTGGCCA	1646
Db	1441	CCGCTGGCACACATTTGGCCTTGGGGAAGGATGAACCCAGCAGTGTGGAGTGAAGTGGCCA	1500
QY	1647	GATGGCAAGATGTGAGCCGGAAAGCTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATC	1706
Db	1501	GATGGCAAGATGTGAGCCGGAAAGCTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATC	1560
QY	1707	CTCTACCCCGGGATGAGGACACATTCAGGACCCAGCCCGGAGTGTGGCAAGGA	1766
Db	1561	CTCTACCCCGGGATGAGGACACATTCAGGACCCAGCCCGGAGTGTGGCAAGGA	1620
QY	1767	TTCTCCCGCAGGAAATATGCGCATTTGATGACACCAATGAATGATCCAGTGTCCCATTC	1826
Db	1621	TTCTCCCGCAGGAAATATGCGCATTTGATGACACCAATGAATGATCCAGTGTCCCATTC	1680
QY	1827	GTGTGCCCTCGAGACAGCCCGTATGTGTCAACACCTATGTAAGCTACAGGTGCCGAGC	1886
Db	1681	GTGTGCCCTCGAGACAGCCCGTATGTGTCAACACCTATGTAAGCTACAGGTGCCGAGC	1740
QY	1887	AACAGAGATGTCAGTCGGGCTACGACCCCAACGAGGATGGCAGAGCTCGTGGG	1942
Db	1741	AACAGAGATGTCAGTCGGGCTACGACCCCAACGAGGATGGCAGAGCTCGTGGG	1796
RESULT 9			
HSA279016			
LOCUS	2507 bp mRNA linear PRI 11-MAR-2001		
DEFINITION	Homo sapiens mRNA for chondrocyte expressed protein 68 kDa (CEP-68 gene).		
ACCESSION	AJ279016		
VERSION	1 GI-10178882		
KEYWORDS	CEP-68 gene; chondrocyte expressed protein 68 kDa.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1		
AUTHORS	Steck,E., Benz,K., Lorenz,H., Loew,M., Gress,T. and Richter,W.		
	Chondrocyte expressed protein-68 (CEP-68), a novel human marker gene for cultured chondrocytes		
JOURNAL	Biochem. J. 353 (Pt 2), 169-174 (2001)		
MEDLINE	21066088		
PUBMED	1119377		
REFERENCE	2. (bases 1 to 2507)		
AUTHORS	Steck,E.		
TITLE	Direct Submission		

JOURNAL Submitted (08-SEP-2000) Steck E., Research, Department of Orthopaedic Surgery, University of Heidelberg, Germany, Schlierbacher Landstr. 200a, Heidelberg, 69118, GERMANY

FEATURES Location/Qualifiers

source 1..2507

organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/cell\_type="chondrocyte"

/tissue\_type="cartilage"

gene 1..2507

/gene="CEP-68"

CDS 1..1962

/gene="CEP-68"

/function="unknown"

/note="protein with EGF-like calcium binding motif"

/codon\_start=1

/product="chondrocyte expressed protein 68 kDa"

/protein\_id="CAC08451.1"

/db\_xref="GI:10178883"

/db\_xref="GOA:Q3NQ79"

/db\_xref="SPTREMBL:Q9NQ79"

/translation="MSRMLPFLLLFLPIETGSQRAEPMFTAVTNSVLPPDVDSNPT QNNAIVGAVTDVHDGDFEIVAGYGNPGLVLVDRAQKLVNIADVERSSPYLALDR QNGAIYACDIDGREGTEIFYLNTNNAFSGVATYDKLFKFRNNEWEDILSDENVVA RVASVAFGRSVACVDRKSGRSYIVIANVAYGVNCGPDALIEMDPRASDLRSGILAR GVDAEAGVSGKVTGGRGVSGYLTILSSASDIFCDNNGPNFLFHNRGDGFVDAAAG VDDPHQGRGVALADNRQKVDIVYGNWNGPHRLYLQMSTHGKVFPRDIASPKFBNP SPVRTVTFADFNQDQLEIFFNNAIYRSSANLFRVIREHGDPLIEELNPGDALER REGATGGVVTFDGDMLDILLSHGESMAQPLSVFRGNQGNFNNWLVVPRTEFGAFA RGAQVLLYTKKSGAHLRIIDGSGYLCEMEPVAHFLGKDGKDEASVEVTPDGMKVRN VASGEMNSVLEILYPRDETLQDPALECGQGFSGQENGCHMDTNECIOFPFVCPDRDK PIVCATYGSVRCRTNKKCSRGVEPNEDTACVGTLGQSFGPRTTTTAAATAAAAAA AQAATPAAPVLVVDGLNLGVSVKESCEPSC"

polyA\_signal 2469..2474

/gene="CEP-68"

/evidence=experimental

ORIGIN

Query Match 80.2%; Score 1745; DB 9; Length 2507;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 147 ATGTCCAGATGTTACCGTTCTCGTCTGCTGCTGCTGCTTCTGCCCATCACTGAGGGTCC 206

Db 1 ATGTCCAGATGTTACCGTTCTGCTGCTGCTGCTTCTGCCCATCACTGAGGGTCC 60

QY 207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCACTCAGTTCTGCCTCTGACTATGAC 266

Db 61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCACTCAGTTCTGCCTCTGACTATGAC 120

QY 267 AGTAATCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGACCATGATGGGAC 326

Db 121 AGTAATCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGACCATGATGGGAC 180

QY 327 TTGTGAGATCGTGTGGCGGGGTACATGACCAACCTGGTCTCTGAAGTATGACCGGGCC 386

Db 181 TTGTGAGATCGTGTGGCGGGGTACATGACCAACCTGGTCTCTGAAGTATGACCGGGCC 240

QY 387 CAGAAGCGGCTGGTGAACATCGCGGTGATGAGCGCAGCTCACCCCTACTACGGCTGCGG 446

Db 241 CAGAAGCGGCTGGTGAACATCGCGGTGATGAGCGCAGCTCACCCCTACTACGGCTGCGG 300

QY 447 GACCGGCAGGGAAACCCATATGGGGTCAAGCGCTGCGACATCGACGGGACGGCCGGAG 506

Db 301 GACCGGCAGGGAAACCCATATGGGGTCAAGCGCTGCGACATCGACGGGACGGCCGGAG 360

QY 507 GAGATCTACTTCTCAACACCAATATGCTTCTCGGGGTGGCCACGTACACCGACAAG 566

Db 361 GAGATCTACTTCTCAACACCAATATGCTTCTCGGGGTGGCCACGTACACCGACAAG 420

QY 567 TTGTTTCAAGTTCCGCAATAAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAAGTGGCC 626

Db 421 TTGTTCAAGTTCCGCAATAACCGTGGGAAGACATCTCTGAGCGATGAGGTCAACGTGGCC 480  
Qy 627 CGTGTGTGGCCAGCCTCTTTGCCGGAAGCTCTGTGGCTGTGTGACAGAAAGGCTCT 686  
Db 481 CGTGTGTGGCCAGCCTCTTTGCCGGAAGCTCTGTGGCTGTGTGACAGAAAGGCTCT 540  
Qy 687 GGAGCCTCTCTATCTACATTCGCAATACGCTACGCTACGCTAAATGTGGCCCTGATGCCCTC 746  
Db 541 GGAGCCTCTCTATCTACATTCGCAATACGCTACGCTAAATGTGGCCCTGATGCCCTC 600  
Qy 747 ATTGAATGGACCCCTGAGCCAGTGAACCTCTCCCGGGCAATCTTGCGCTCAGAGATGTG 806  
Db 601 ATTGAATGGACCCCTGAGCCAGTGAACCTCTCCCGGGCAATCTTGCGCTCAGAGATGTG 660  
Qy 807 GCTGCTGAGGCTGGGTGAGCAATATACAGGGGGCCGAGGCGTACGCTGGGGCCCATC 866  
Db 661 GCTGCTGAGGCTGGGTGAGCAATATACAGGGGGCCGAGGCGTACGCTGGGGCCCATC 720  
Qy 867 CTCAGCAGCAGTGCCTCGGATATCTTTCGCAATAGAAATGGGCTAACTTCTTTTC 926  
Db 721 CTCAGCAGCAGTGCCTCGGATATCTTTCGCAATAGAAATGGGCTAACTTCTTTTC 780  
Qy 927 CACAAACCGGGCGATGGCACTTTGTGGACGCTGGGGCCAGTGTGTGAGCAACCC 986  
Db 781 CACAAACCGGGCGATGGCACTTTGTGGACGCTGGGGCCAGTGTGTGAGCAACCC 840  
Qy 987 CACCAAGCATGGGAGGTGTGCGCTGCTGCTTCAACCGTGTGGCAAGTGCACATC 1046  
Db 841 CACCAAGCATGGGAGGTGTGCGCTGCTGCTTCAACCGTGTGGCAAGTGCACATC 900  
Qy 1047 GTCTATGGCAATGGCAATGGGCGCCACCGCTCTATCTGCAATAGCAACCCATGGGAAG 1106  
Db 901 GTCTATGGCAATGGCAATGGGCGCCACCGCTCTATCTGCAATAGCAACCCATGGGAAG 960  
Qy 1107 GTCCGCTTCCGGGACATCGCCTCAACCAAGTTCTCCATGCCCTCCCTCTCGCAAGTTC 1166  
Db 961 GTCCGCTTCCGGGACATCGCCTCAACCAAGTTCTCCATGCCCTCCCTCTCGCAAGTTC 1020  
Qy 1167 ATCCCGCGGCTTTGACAAATACAGGAGCTGGAGATCTTCTCAACAACTGCGCTAC 1226  
Db 1021 ATCCCGCGGCTTTGACAAATACAGGAGCTGGAGATCTTCTCAACAACTGCGCTAC 1080  
Qy 1227 CGCAGCTCTCAGCAACCGCTCTTCCGCGTATCCGTAGAGACGAGACCCCTTC 1286  
Db 1081 CGCAGCTCTCAGCAACCGCTCTTCCGCGTATCCGTAGAGACGAGACCCCTTC 1140  
Qy 1287 ATCGAGGAGCTCAATCCCGGCGACGCTTTGGAGCTGAGGGCGGGGCAAGGGGTGTG 1346  
Db 1141 ATCGAGGAGCTCAATCCCGGCGACGCTTTGGAGCTGAGGGCGGGGCAAGGGGTGTG 1200  
Qy 1347 GTGACCGCTTCGAGGAGCGGATGCTGGACCTCATCTGTCCATGGAGATCCATG 1406  
Db 1201 GTGACCGCTTCGAGGAGCGGATGCTGGACCTCATCTGTCCATGGAGATCCATG 1260  
Qy 1407 GCTCAGCCGCTCTCGCTTTCGGGGCAATCAGGGCTTCAACAACTGCTGCGAGTG 1466  
Db 1261 GCTCAGCCGCTCTCGCTTTCGGGGCAATCAGGGCTTCAACAACTGCTGCGAGTG 1320  
Qy 1467 GTGCCACGACCGGTTTGGGGCTTTGGCGAGCTGAGGCTAAGTGTGTCTACACCAAG 1526  
Db 1321 GTGCCACGACCGGTTTGGGGCTTTGGCGAGCTGAGGCTAAGTGTGTCTACACCAAG 1380  
Qy 1527 AAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTTACCTGTGTGAGATGGAG 1586  
Db 1381 AAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTTACCTGTGTGAGATGGAG 1440  
Qy 1587 CCGGTGGCAACATTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGAGAGTGAAGTGGCCA 1646  
Db 1441 CCGGTGGCAACATTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGAGAGTGAAGTGGCCA 1500  
Qy 1647 GATGGCAAGATGGTGAAGCGGACGTGGCCAGCGGGGAGATGAATCACTCACTGTGGAGATC 1706  
Db 1501 GATGGCAAGATGGTGAAGCGGGAACGTGGCCAGCGGGGAGATGAATCACTCACTGTGGAGATC 1560

Qy 1707 CTCTACCCCGGATGAGGACACACTTCAGGACCCAGGCCCCACTGGAGTGTGGCCAAAGA 1766  
Db 1561 CTCTACCCCGGATGAGGACACACTTCAGGACCCAGGCCCCACTGGAGTGTGGCCAAAGA 1620  
Qy 1767 TTCTCCAGCAGCAAAATGGCCATTCATGGAACCAATGAATGATCCAGTTCCTCATTC 1826  
Db 1621 TTCTCCAGCAGCAAAATGGCCATTCATGGAACCAATGAATGATCCAGTTCCTCATTC 1680  
Qy 1827 GTGTCCCTCGAGACAGCCCGTATGTCAACACTATGGAAGCTTACAGTTCGGGACC 1886  
Db 1681 GTGTCCCTCGAGACAGCCCGTATGTCAACACTATGGAAGCTTACAGTTCGGGACC 1740  
Qy 1887 AACCAAGAGTGCAGTCCGGGCTACGAGCCCAACGAGGATGGCACAGCTTCGCTGGG 1942  
Db 1741 AACCAAGAGTGCAGTCCGGGCTACGAGCCCAACGAGGATGGCACAGCTTCGCTGGG 1796  
RESULT 10  
AX470056  
LOCUS AX470056 1947 bp DNA linear PAT 09-AUG-2002  
DEFINITION Sequence 22 from Patent WO02053709.  
ACCESSION AX470056  
VERSION AX470056.1 GI:22205301  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
Richter, W. and Steck, E.  
TITLE Cartilage marker gene, method for the production and use thereof  
JOURNAL Patent: WO 02053709-A 22 11-JUL-2002;  
CYTONET GMBH & CO KG (DE)  
FEATURES  
source  
1. .1947  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 80.1%; Score 1744; DB 6; Length 1947;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 147 ATGTCCAGGATG-TACCGTTCTCTGCTGCTGTGTTTCTGCCATCACTGAGGGTCC 206  
Db 1 ATGTCCAGGATG-TACCGTTCTCTGCTGCTGTGTTTCTGCCATCACTGAGGGTCC 60  
Qy 207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCACTCAGTTCTGCTCTGACTATGAC 266  
Db 61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCACTCAGTTCTGCTCTGACTATGAC 120  
Qy 267 AGTAATCCACAGCTCAACTATGTTGGGAGTTACTGATGTGACCATGATGAGGGAC 326  
Db 121 AGTAATCCACAGCTCAACTATGTTGGGAGTTACTGATGTGACCATGATGAGGGAC 180  
Qy 327 TTTGAGATCGTGTGGGGGTACAACTGACCCCACTGTTCTTCTGAGTATGACCGGGCC 386  
Db 181 TTTGAGATCGTGTGGGGGTACAACTGACCCCACTGTTCTTCTGAGTATGACCGGGCC 240  
Qy 387 CAGAAGCGGCTGTGAACATCGCGGTTCGATGAGCGCAGCTACCCCTTACTACGCGTGGG 446  
Db 241 CAGAAGCGGCTGTGAACATCGCGGTTCGATGAGCGCAGCTACCCCTTACTACGCGTGGG 300  
Qy 447 GACCGGAGGGGAAAGCCATTTGGGGTACAGCTCGGACATCGACGGGACGGCCGGGAG 506  
Db 301 GACCGGAGGGGAAAGCCATCGGGGTTCACAGCTCGGACATCGACGGGACGGCCGGGAG 360  
Qy 507 GAGATCTACTTCTCAACACCAATATGCTTCTCGGGGGTGGCCACGTTACACCGACAAG 566  
Db 361 GAGATCTACTTCTCAACACCAATATGCTTCTCGGGGGTGGCCACGTTACACCGACAAG 420

```
QY 567 TTGTTCAAGTTCGCGAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAAGTGGCC 626
Db 421 TTGTTCAAGTTCGCGAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAAGTGGCC 480
QY 627 CGTGGTGTGGCCAGCCTCTTTTGCAGACGCTCTGTGTGCCCTGTGTGGACAGAAAGGCTCT 686
Db 481 CGTGGTGTGGCCAGCCTCTTTTGCAGACGCTCTGTGTGCCCTGTGTGGACAGAAAGGCTCT 540
QY 687 GGAAGTACTCTATCTACATTTGCAATTAACGCTTACGCTTACGCTTAAATGTGGCCCTGTATGCCCTC 746
Db 541 GGAAGTACTCTATCTACATTTGCAATTAACGCTTACGCTTAAATGTGGCCCTGTATGCCCTC 600
QY 747 ATTGAATGACCTTGAAGCAGTACCTCTCCGGGGCAATTCTGGCGCTCAGAGATGTG 806
Db 601 ATTGAATGACCTTGAAGCAGTACCTCTCCGGGGCAATTCTGGCGCTCAGAGATGTG 660
QY 807 GCTGCTGAGCTGGGCTCAGCAATATACAGGGGGCGAGGCGTTCAGCGTGGGCCCCCATC 866
Db 661 GCTGCTGAGCTGGGCTCAGCAATATACAGGGGGCGAGGCGTTCAGCGTGGGCCCCCATC 720
QY 867 CTCAGACAGTGCCTCGGATATCTTCTGCGACAATGAGAAATGGGCTAACTTCTCTTTTC 926
Db 721 CTCAGACAGTGCCTCGGATATCTTCTGCGACAATGAGAAATGGGCTAACTTCTCTTTTC 780
QY 927 CACAACGGGGCATGGCACCTTTGTGGACGCTCGGGCCAGTCTGTGTGGAGAGACCC 986
Db 781 CACAACGGGGCATGGCACCTTTGTGGACGCTCGGGCCAGTCTGTGTGGAGAGACCC 840
QY 987 CACGAGTGGGCGAGGTGTGCGCTTGGCTGACTTCAACCGTGTATGCAAGTGGACATC 1046
Db 841 CACGAGTGGGCGAGGTGTGCGCTTGGCTGACTTCAACCGTGTATGCAAGTGGACATC 900
QY 1047 GTCTATGGCAATGGAATGGCCCCCACCCTCTATCTGCAATGAGCACCCCATGGGAAG 1106
Db 901 GTCTATGGCAATGGAATGGCCCCCACCCTCTATCTGCAATGAGCACCCCATGGGAAG 960
QY 1107 GTCCGCTTCGGGACATCGCTCACCAAGTCTCCATGCGCTCCCTGTGCCACGCTC 1166
Db 961 GTCCGCTTCGGGACATCGCTCACCAAGTCTCCATGCGCTCCCTGTGCCACGCTC 1020
QY 1167 ATCACCAGCTCTTACCAATGACAGGAGCTGGAGATCTTCTTCAACACATTTGCTTAC 1226
Db 1021 ATCACCAGCTCTTACCAATGACAGGAGCTGGAGATCTTCTTCAACACATTTGCTTAC 1080
QY 1227 CGAGCTCTCTACCAACCGCTCTTCCGCGTCAATCCGTAGAGACACGAGACCCCTC 1286
Db 1081 CGAGCTCTCTACCAACCGCTCTTCCGCGTCAATCCGTAGAGACACGAGACCCCTC 1140
QY 1287 ATCGAGGAGCTCAATCCCGGCGAGCGCTTGGAGCTTGAGGCGCGGGCCACAGGGGTGTG 1346
Db 1141 ATCGAGGAGCTCAATCCCGGCGAGCGCTTGGAGCTTGAGGCGCGGGCCACAGGGGTGTG 1200
QY 1347 GTGACCGACTTTCAGCGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db 1201 GTGACCGACTTTCAGCGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260
QY 1407 GCTCAGCGGTGTCGCTTTCGGGGCAATCAGGGTCTCAACCAACCTCGCTGCGAGTG 1466
Db 1261 GCTCAGCGGTGTCGCTTTCGGGGCAATCAGGGTCTCAACCAACCTCGCTGCGAGTG 1320
QY 1467 GTCCACGACACCGGTTTGGGCTTTGGCGCTTGGCAGGGAGCTTAAGTCTGCTCTACCAAG 1526
Db 1321 GTCCACGACACCGGTTTGGGCTTTGGCGCTTGGCAGGGAGCTTAAGTCTGCTCTACCAAG 1380
QY 1527 AAGAGTGGGCCCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db 1381 AAGAGTGGGCCCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440
QY 1587 CCGTGGCACACTTTGGCTTGGGAAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCCA 1646
Db 1441 CCGTGGCACACTTTGGCTTGGGAAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCCA 1500
QY 1647 GATGGCAAGATGTGAGCCGGAACGTGGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATC 1706
```

```
Db 1501 GATGGCAAGATGTGAGCCGAAACGTGGCCAGCGGGAGATGAACCTCAGTCTGGAGATC 1560
QY 1707 CTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGA 1766
Db 1561 CTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGA 1620
QY 1767 TTCTCCAGCAGGAAATGSCCATTCGATGAGCACCAATGAATGCAATTCAGTTCCTCCATTC 1826
Db 1621 TTCTCCAGCAGGAAATGSCCATTCGATGAGCACCAATGAATGCAATTCAGTTCCTCCATTC 1680
QY 1827 GTGTGCTTCGAGACAGACCCGATGTGTGTCAACACCTATGGAAGCTACAGTGGCCGAC 1886
Db 1681 GTGTGCTTCGAGACAGACCCGATGTGTGTCAACACCTATGGAAGCTACAGTGGCCGAC 1740
QY 1887 AACAGAAGTGCAGTTCGGGCTACGAGCCCAACGAGGATGGCACAGCTCCGTGG 1941
Db 1741 AACAGAAGTGCAGTTCGGGCTACGAGCCCAACGAGGATGGCACAGCTCCGTGG 1795

RESULT 11
HSA421515
LOCUS HSA421515 1947 bp mRNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens mRNA for CRTAC1-B protein (CRTAC1 gene).
ACCESSION AJ421515
VERSION AJ421515.1 GI:19171210
KEYWORDS CRTAC1 gene; CRTAC1-B protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Steck,E. and Richter,W.
TITLE A novel form of CRTAC1 expressed in human and mouse brain generated by alternative splicing of a newly identified last exon
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1947)
AUTHORS Steck,E.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2001) Steck E., Research, Department of Orthopaedic Surgery, University of Heidelberg, Schlierbacher Landstr. 200a, Heidelberg, 69118, GERMANY
FEATURES
source
1..1947
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="brain"
1..1947
/genes="CRTAC1"
/genes="CRTAC1"
/codon_start=1
/evidence=experimental
/product="CRTAC1-B protein"
/protein_id="CAD13394.1"
/db_xref="GI:19171211"
/db_xref="GOA:O8TE52"
/db_xref="SPTREMBL:O8TE52"
/translation="MSRMLPFLLLWFLPITGSGRAEPMFTAVTNSVLPPDYDSNPT
QLNYGVAVTDVDDHGDGFETVAGYGNPLVLKYDRAQKLVNIAVDRESPPYALDR
QNALIGVTACIDGDGREIYFLNTNNAFSGVATYTDKLFKRNKRWEDILSDVNVA
RGVASLFAGRSVACVDKRGSGRSYIVANYGVNVPDALIEMDPASDLSCILLAR
DYAAEAGVSKYTGRCVSVPILSSASDIFCDNENGPNFLQMSHGKVRFDRIASPFSNP
VDDPQHGRGVALADPNRDGKVDIVGNWNGPHRLYLQMSHGKVRFDRIASPFSNP
SVPTVTITADFNDQLELFFNNIAYRSSANRLFRVIRREHGDPLIEELNPGDALEP
EERGQGVVYTFDFGMDLILSHGSEMAQPLSVFRGNQGFNNMLRVVYTRTFGFAF
RAKVVLVYTKSGAHLRIIDGSGYLICEMEPVAHFLGKDEASVEVTPDGMVSRN
VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSGQENGHCMDTNECIQFPFVCPDRDK
PVCVNTYGSYRCRTNKKCSRGVEPNEDGTACVAQVAFLGYSYSAASRISEPLRSALY
SLGLGLQLQLYAL"
1923..1928
polyA_signal
```



Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,  
Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,  
Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Wagatsuna,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2889)  
Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

## FEATURES

source

Location/Qualifiers

1..2889  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="SYNOV1000118"  
/tissue\_type="synovial membrane tissue from rheumatoid  
arthritis"  
/clone\_lib="SYNOV1"  
/notes="cloning vector: pME18SFL3"

## ORIGIN

Query Match 77.5%; Score 1687; DB 9; Length 2889;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1937; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CGGAGGCTCGAGGCCCGGACCGGGCTGGGAGCAAGCAGCGGGCGCGCGG 60  
DB 235 CGGAGGCTCGAGGCCCGGACCGGGCTGGGAGCAAGCAGCGGGCGCGCGG 294  
QY 61 GCAGGCGGCGACGAGCGCGCGCTTCCACGCGCCCTAGCGCGGGCGGAGCGGGA 120  
DB 295 GCAGGCGGCGACGAGCGCGCGCTTCCACGCGCCCTAGCGCGGGCGGAGCGGGA 354  
QY 121 GGATGGCTCGAGCGCTGACCCGGCATGTCAGGATGTTACGGTTCCTGCTGCTCT 180  
DB 355 GGATGGCTCGAGCGCTGACCCGGCATGTCAGGATGTTACGGTTCCTGCTGCTCT 414  
QY 181 GGTTTCTGCCATCACTGAGGGTCCAGCGGCTGAAACCCATGTTCACTGCAAGTCACCA 240  
DB 415 GGTTTCTGCCATCACTGAGGGTCCAGCGGCTGAAACCCATGTTCACTGCAAGTCACCA 474  
QY 241 ACTCAGTTCTGCTCTGATGACAGTAATCCACCCAGCTCAACTATGTTGGCGAG 300  
DB 475 ACTCAGTTCTGCTCTGATGACAGTAATCCACCCAGCTCAACTATGTTGGCGAG 534  
QY 301 TTACTGATGTCGACCATGATGGGACTTTGAGATCGTGGGGGGGTACAATGGACCCA 360  
DB 535 TTACTGATGTCGACCATGATGGGACTTTGAGATCGTGGGGGGGTACAATGGACCCA 594  
QY 361 ACCTGGTTCTGAAGTATGACCGGGCCAGAAAGCGGCTGGTGAACATCGCGGTGATGAGC 420  
DB 595 ACCTGGTTCTGAAGTATGACCGGGCCAGAAAGCGGCTGGTGAACATCGCGGTGATGAGC 654  
QY 421 GCAGCTCACCTACTACGGCTCGGGAACCGGAGGGGAAACCGCATTTGGGGTCAAGCCT 480  
DB 655 GCAGCTCACCTACTACGGCTCGGGAACCGGAGGGGAAACCGCATTTGGGGTCAAGCCT 714  
QY 481 GCGACATCGAGGGGAGCGGGAGAGATCTACTTCCTCAACACCAATAATGCTTCT 540

DB 715 GCGACATCGACGCGGAGCGCGGAGGAGATCTACTTCTCAACACCAATAATGCTTCT 774  
QY 541 CGGGGGTGGCCACGTCACCGCAAGTGTTCAGTTCGCAATTAACCGGTGGGAAGACA 600  
DB 775 CGGGGGTGGCCACGTCACCGCAAGTGTTCAGTTCGCAATTAACCGGTGGGAAGACA 834  
QY 601 TCCTGAGCGATGAGGTCAACGTCGCGCGTGGTGGCGAGCCTCTTTGCCGGAGCGCTCTG 660  
DB 835 TCCTGAGCGATGAGGTCAACGTCGCGCGTGGTGGCGAGCCTCTTTGCCGGAGCGCTCTG 894  
QY 661 TGGCCTGTGTGAGCAAGAGGCTCTGGAGCTACTCTATCTACATTCGCAATTAACGCT 720  
DB 895 TGGCCTGTGTGAGCAAGAGGCTCTGGAGCTACTCTATCTACATTCGCAATTAACGCT 954  
QY 721 ACGGTAATGTGGCCCTGATGCCCTCATTTGAAATGACCGCTGAGGCGAGTGACCTCTCC 780  
DB 955 ACGGTAATGTGGCCCTGATGCCCTCATTTGAAATGACCGCTGAGGCGAGTGACCTCTCC 1014  
QY 781 GGGGCAATTCGGCGCTCAGAGATGTGGTCTGAGGCTGGGGTCAAGCAATATACAGGGG 840  
DB 1015 GGGGCAATTCGGCGCTCAGAGATGTGGTCTGAGGCTGGGGTCAAGCAATATACAGGGG 1074  
QY 841 GCGGAGCGCTCAGCGTGGGCGCCATCTCTCAGCAGAGTGCCTCGATATCTTCTGGACA 900  
DB 1075 GCGGAGCGCTCAGTGTGGGCGCCATCTCTCAGCAGAGTGCCTCGATATCTTCTGGACA 1134  
QY 901 ATGAGAATGGGCTAACTTCTTTCCACAACCGGGCGATGACCTTTGTGGAGCGCTG 960  
DB 1135 ATGAGAATGGGCTAACTTCTTTCCACAACCGGGCGATGACCTTTGTGGAGCGCTG 1194  
QY 961 CGGCCAGTGTGTGTGGAGCACCCCGCACAGCATGGCGAGGTGTGCGCCTGGCTGACT 1020  
DB 1195 CGGCCAGTGTGTGTGGAGCACCCCGCACAGCATGGCGAGGTGTGCGCCTGGCTGACT 1254  
QY 1021 TCAACCGTATGGCAAGTGGACATCGTCTATGGCACTGGAAATGGCCCCCGCCTCT 1080  
DB 1255 TCAACCGTATGGCAAGTGGACATCGTCTATGGCACTGGAAATGGCCCCCGCCTCT 1314  
QY 1081 ATCTGCAATAGACACCCATGGAGGTCCGCTTCGGGACATCGCTCACCAAGTTCT 1140  
DB 1315 ATCTGCAATAGACACCCATGGAGGTCCGCTTCGGGACATCGCTCACCAAGTTCT 1374  
QY 1141 CCATGCCCCCTCCCTGTCCGACAGGTCAACCGCCGACCTTTGACAAATGACAGGAGCTGG 1200  
DB 1375 CCATGCCCCCTCCCTGTCCGACAGGTCAACCGCCGACCTTTGACAAATGACAGGAGCTGG 1434  
QY 1201 AGATCTTCTTCAACACATGCTACCGAGCTCTCAGCCAAACCGCTCTTCGGGTCA 1260  
DB 1435 AGATCTTCTTCAACACATGCTACCGAGCTCTCAGCCAAACCGCTCTTCGGGTCA 1494  
QY 1261 TCCGTAGAGACGAGAGACCCCTCATCGAGAGTCAATCCCGCGAGCGCTTGGAGC 1320  
DB 1495 TCCGTAGAGACGAGAGACCCCTCATCGAGAGTCAATCCCGCGAGCGCTTGGAGC 1554  
QY 1321 CTGAGGCGCGGGCAACAGGGGGTGTGGTACCGACTTCGACGAGAGCGGATGCTGGACC 1380  
DB 1555 CTGAGGCGCGGGCAACAGGGGGTGTGGTACCGACTTCGACGAGAGCGGATGCTGGACC 1614  
QY 1381 TCATCTTGTCCCATGAGAGTCAATGGCTCAGCGCTGTCCGCTTCGCGGGCAATCAGG 1440  
DB 1615 TCATCTTGTCCCATGAGAGTCAATGGCTCAGCGCTGTCCGCTTCGCGGGCAATCAGG 1674  
QY 1441 GCTTCAACAACTGGCTGCGAGTGGTCCACGACCGCGTTTGGGGCTTTGGCGAGG 1500  
DB 1675 GCTTCAACAACTGGCTGCGAGTGGTCCACGACCGGGTTTGGGGCTTTGGCGAGG 1734  
QY 1501 GAGCTAAGTCTGTCTTACCAAGAGAGTGGGGCCCACTGAGGATCATCGACGGG 1560  
DB 1735 GAGCTAAGTCTGTCTTACCAAGAGAGTGGGGCCCACTGAGGATCATCGACGGG 1794  
QY 1561 GCTCAGGCTACCTGTGTGAGTGGAGCGCGTGGCACACTTTGGCTGGGAGGATGAAG 1620





QY	1339	GGGTTGTTGGTACCGACTTCACGAGACGGGATGCTGGACCTCATCTTGTGCCATGAG	1398
Db	1638	GGGTTGTTGGTACCGACTTCACGAGACGGGATGCTGGACCTCATCTTGTGCCATGAG	1697
QY	1399	AGTCCATGGCTACCGCTGTCCTCTTCGGGGCAATCAGGGCTCAACACAACTGGC	1458
Db	1698	AGTCCATGGCTACCGCTGTCCTCTTCGGGGCAATCAGGGCTCAACACAACTGGC	1757
QY	1459	TGGAGTGGTGCCACGACCCGGTTTGGGGCTTTTGGCAGGGAGCTAAAGTCTGCTCT	1518
Db	1758	TGGAGTGGTGCCACGACCCGGTTTGGGGCTTTTGGCAGGGAGCTAAAGTCTGCTCT	1817
QY	1519	ACACCAAGAAGTGGGGCCACCTAGGATCATCGACGGGGCTCAGGCTACTGTGTG	1578
Db	1818	ACACCAAGAAGTGGGGCCACCTAGGATCATCGACGGGGCTCAGGCTACTGTGTG	1877
QY	1579	AGATGGAGCCCGTGGCACACTTTGGCTGGGGAAGGATGAAGCCAGCAGTGTGGAGTGA	1638
Db	1878	AGATGGAGCCCGTGGCACACTTTGGCTGGGGAAGGATGAAGCCAGCAGTGTGGAGTGA	1937
QY	1639	CGTGGCCAGATGGCAAGATGCTGAGCCGGAACGTTGGCCAGCGGGAGATGAATCACTGTC	1698
Db	1938	CGTGGCCAGATGGCAAGATGCTGAGCCGGAACGTTGGCCAGCGGGAGATGAATCACTGTC	1997
QY	1699	TGGAGATCTCTTACCCCGGGATGAGACACACTTCAGGACCCAGCCCTGAGTGTG	1758
Db	1998	TGGAGATCTCTTACCCCGGGATGAGACACACTTCAGGACCCAGCCCTGAGTGTG	2057
QY	1759	GCCAAAGATCTCCACGACGAAATGGCCATTCATGGACACCAATGAATGCATCCAGT	1818
Db	2058	GCCAAAGATCTCCACGACGAAATGGCCATTCATGGACACCAATGAATGCATCCAGT	2117
QY	1819	TCCATTGCTGTCCTCGAGACAGCCCGTATGTGTCAACACTATGGAAGCTACAGGT	1878
Db	2118	TCCATTGCTGTCCTCGAGACAGCCCGTATGTGTCAACACTATGGAAGCTACAGGT	2177
QY	1879	GCCGGACCAACAGATGTCAGTCCGGGCTACAGCCCAAGGATGGACAGCCTGGC	1938
Db	2178	GCCGGACCAACAGATGTCAGTCCGGGCTACAGCCCAAGGATGGACAGCCTGGC	2237
QY	1939	TGG 1941	
Db	2238	TGG 2240	
RESULT 14			
BC034245			
LOCUS	2147 bp	mrna	linear PRI 03-OCT-2003
DEFINITION	Homo sapiens cartilage acidic protein 1, mRNA (cdna clone MGC:16989		
IMAGE:4178363			
ACCESSION	BC034245		
VERSION	BC034245.1	GI:21706427	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 2147)		
	Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,		
	Strausberg, R.L., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D.,		
	Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,		
	Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,		
	Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,		
	Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavani, T.L.,		
	Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,		
	Carninci, P., Prance, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,		
	Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,		
	McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,		
	Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,		
	Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,		
	Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,		
	Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,		

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 2147)  
Strausberg, R.  
Direct Submission  
Submitted (02-JUL-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-x@mail.nih.gov](mailto:cgaps-x@mail.nih.gov)  
Tissue Procurement: David N. Louis, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunsatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 12 Row: h Column: 5  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 8922351.

## FEATURES

source  
1. .2147  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:16989 IMAGE:4178363"  
/tissue\_type="Brain, anaplastic oligodendroglioma with  
1p/19q loss"  
/clone\_lib="NCI CGAP\_Brn67"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. .2147  
/genes="CRTAC1"  
/notes="synonyms: FLJ10320, CEP-69, ASPIC1"  
/db\_xref="LocusID:55118"  
/db\_xref="MIM:606276"  
249\_-1604  
/codon\_start=1  
/product="CRTAC1 protein"  
/protein\_id="AAH34245.1"  
/db\_xref="GI:21706428"  
/db\_xref="LocusID:55118"  
translation="MDPEASDLSRGILLRDVAEAGVSKYTGGRGVSGVPILSSAS  
DIFCDNENGFNLFHNRGGTGVDAASAGVDDPHOGRGVALADFNDRGKVDIVYGN  
WNGPHRLYLQMSHGKVRDEEINSPKSPVPTVITADFNQDELEIFNNIAYRS  
SSANLFRVIRREHGDPLEEDNPGDALEPGRTGGVTVDFDGMGLDILSHGSM  
AQLSVFRGNQFNNNLRVPRFGAFARGAKVLYTKGSAHLRIIDGSGYLCE  
MEPVAHFGLGKDEASSVEVTPDGKMSRVNAGSMNSVLEILYPRDETLQDPAPLE  
CGQGSQOENHCHMTNECIQPFVCPKPVCVNTYGSYRCRTNKKCRGTEPNEDG  
TACVTLGSGSPGPRPTPTTAAATAAAAAAGATAAPVLVDGDLNIGSVKESCEPS  
C"  
1293..1436  
/note="EGF CA; Region: Calcium-binding EGF-like domain"  
/db\_xref="CDD:smart00179"

## misc\_feature

## ORIGIN

Query Match

66.1%; Score 1438; DB 9; Length 2147;

Best Local Similarity 100.0%; Pred. No. 0;		Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	505	AGGAGATCTACTTCTCTCAACCAATAATATGCTTCTCGGGGTGTCACGATACACGACA	564
Db	1	AGGAGATCTACTTCTCTCAACCAATAATATGCTTCTCGGGGTGTCACGATACACGACA	60
QY	565	AGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCTGACGATGAGGTCAAGTGG	624
Db	61	AGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCTGACGATGAGGTCAAGTGG	120
QY	625	CCGCTGGTGTGCGACGCTCTTTGCGGACGCTCTGTGGCCCTGTGTGGACAGAAAGGCT	684
Db	121	CCGCTGGTGTGCGACGCTCTTTGCGGACGCTCTGTGGCCCTGTGTGGACAGAAAGGCT	180
QY	685	CTGACGCTTACTTACTTACTATTCGCAATTACGCTTACGGTATGAGGCTCTGATGCC	744
Db	181	CTGACGCTTACTTACTTACTATTCGCAATTACGCTTACGGTATGAGGCTCTGATGCC	240
QY	745	TCATTGMAATGACCTTGAGGCACTGACCTCTCCGGGGCATTTCTGGCGCTCAGAGATG	804
Db	241	TCATTGMAATGACCTTGAGGCACTGACCTCTCCGGGGCATTTCTGGCGCTCAGAGATG	300
QY	805	TGGCTGCTGAGGCTGGGTCAGCAATAATACAGGGGGCCGAGCGCTCAGCTGGGCCCCA	864
Db	301	TGGCTGCTGAGGCTGGGTCAGCAATAATACAGGGGGCCGAGCGCTCAGCTGGGCCCCA	360
QY	865	TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATAGAAATGGGCTTAATCTCTTT	924
Db	361	TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATAGAAATGGGCTTAATCTCTTT	420
QY	925	TCACAAACCGGGCGATGACCTTTGTGAGCGCTGCGGCCAGTCTGCTGTGGAGACC	984
Db	421	TCACAAACCGGGCGATGACCTTTGTGAGCGCTGCGGCCAGTCTGCTGTGGAGACC	480
QY	985	CCCACAGCATGGGGGAGGTGCGCCCTGGCTGACTTCAACCGTGTATGGCAAAAGTGGACA	1044
Db	481	CCCACAGCATGGGGGAGGTGCGCCCTGGCTGACTTCAACCGTGTATGGCAAAAGTGGACA	540
QY	1045	TGCTTATGGAATGGAATGGCCGCCACCGCTCTATCTGCAATAGACACCCATGGGA	1104
Db	541	TGCTTATGGAATGGAATGGCCGCCACCGCTCTATCTGCAATAGACACCCATGGGA	600
QY	1105	AGTTCGCTTCGGGACATGCGCTCACCAAGTTCCTCATGCTCCCTGTCGCAACG	1164
Db	601	AGTTCGCTTCGGGACATGCGCTCACCAAGTTCCTCATGCTCCCTGTCGCAACG	660
QY	1165	TCATCACCCGCACTTTGCAATGACACGAGCTGGAGATCTTCTTCAACCAATTCGCT	1224
Db	661	TCATCACCCGCACTTTGCAATGACACGAGCTGGAGATCTTCTTCAACCAATTCGCT	720
QY	1225	ACGCGAGCTCTCAGCAACCGCTTTCGGGTGATCGTAGAGACGAGACCCCC	1284
Db	721	ACGCGAGCTCTCAGCAACCGCTTTCGGGTGATCGTAGAGACGAGACCCCC	780
QY	1285	TCATCAGGAGGTCAATCCGCGACGCTTGAGGCTTCAGGCGCGGGCACAGGGGGTG	1344
Db	781	TCATCAGGAGGTCAATCCGCGACGCTTGAGGCTTCAGGCGCGGGCACAGGGGGTG	840
QY	1345	TGTTGACCGCACTTCGACGAGAGCGGATGCTGACCTCATCTTGTCCCATGGAGATCCA	1404
Db	841	TGTTGACCGCACTTCGACGAGAGCGGATGCTGACCTCATCTTGTCCCATGGAGATCCA	900
QY	1405	TGGCTCAGCGCTGTCTGCTTTCGGGGCAATCAGGGCTTCAACCAACTGCTGGAG	1464
Db	901	TGGCTCAGCGCTGTCTGCTTTCGGGGCAATCAGGGCTTCAACCAACTGCTGGAG	960
QY	1465	TGTTGCAACGACCCGCTTTGGGGCTTTGCCAGGGGAGCTAAGGTCTGTCTACACCA	1524
Db	961	TGTTGCAACGACCCGCTTTGGGGCTTTGCCAGGGGAGCTAAGGTCTGTCTACACCA	1020
QY	1525	AGAAGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGG	1584
Db			

Db	1021	AGAACAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGG	1080
QY	1585	AGCCCGTGGCACACTTTGCGCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC	1644
Db	1081	AGCCCGTGGCACACTTTGCGCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC	1140
QY	1645	CAGATGGCAAGATGCTGAGCCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGA	1704
Db	1141	CAGATGGCAAGATGCTGAGCCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGA	1200
QY	1705	TCCTCTAACCCCGGATGAGACACATTCAGAGACCCAGCCCACTGGAGTGTGGCCAAG	1764
Db	1201	TCCTCTAACCCCGGATGAGACACATTCAGAGACCCAGCCCACTGGAGTGTGGCCAAG	1260
QY	1765	GATTCCTCCAGCAGAAATGCCATTGCATGACACCAATGAATGATCCAGTTCCTCAT	1824
Db	1261	GATTCCTCCAGCAGAAATGCCATTGCATGACACCAATGAATGATCCAGTTCCTCAT	1320
QY	1825	TGCTGTGCCCTCAGACAAAGCCCGTATGTGTCAACACCTATGAAAGCTACAGTGTCCGA	1884
Db	1321	TGCTGTGCCCTCAGACAAAGCCCGTATGTGTCAACACCTATGAAAGCTACAGTGTCCGA	1380
QY	1885	CCAAACAAGAGTGCATGCGGGCTACGAGCCCAACAGAGATGGCACAGCTCGGTGGG	1942
Db	1381	CCAAACAAGAGTGCATGCGGGCTACGAGCCCAACAGAGATGGCACAGCTCGGTGGG	1438
RESULT 15			
AX202100		2263 bp	DNA
LOCUS		Sequence 30 from Patent WO0153531.	linear PAT 30-AUG-2001
DEFINITION		AX202100	
ACCESSION		AX202100	
VERSION		AX202100.1	GI:15391887
KEYWORDS			
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		1. Phippard, D., Vasanthakumari, G., Dorson, S. and Ma, X.J.	
AUTHORS		Osteoarthritis tissue derived nucleic acids, polypeptides, vectors,	
TITLE		Patent: WO 0153531-A 30 26-JUL-2001;	
JOURNAL		Pharmacia Corporation (US)	
FEATURES		Location/Qualifiers	
source		1..2263	
/organism="Homo sapiens"			
/mol_type="unassigned DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Query Match		65.6%;	Score 1428; DB 6; Length 2263;
Best Local Similarity		99.6%;	Pred. No. 0;
Matches 1918;		Conservative 0;	Mismatches 5; Indels 2; Gaps 2;
QY	19	CCCGGACCGGGCTGGAGCAAGCAGCGGGCGGGCGCGGCGGCGGCGGCGGCGGCGGCGG	78
Db	153	CCCGGACCGGGCTGGAGCAAGCAGCGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGG	212
QY	79	GCCCGTTCGACCGCCCTTAGCGCGGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG	138
Db	213	GCCCGTTCGACCGCCCTTAGCGCGGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG	272
QY	139	ACCCGGCATGTCAGAGTGTACCGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	198
Db	273	ACCCGGCATGTCAGAGTGTACCGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	332
QY	199	AGGGGTCCGAGCGGGCTGAACCCATGTTCACTGCACTACCAACTCAGTTCCTGCTGCTG	258
Db	333	AGGGGTCCGAGCGGGCTGAACCCATGTTCACTGCACTACCAACTCAGTTCCTGCTGCTG	392
QY	259	ACTATGACAGTAAATCCACCCAGCTCAACTATGCTGGCAGTTACTGATGTGACCATG	318
Db			



Db 393 ACTATGACAGTAAATCCACCCAGCTCAACTATGCTGGCAGTTACTGATGTGGACCAATG 452  
QY 319 ATGGGACTTTGAGATCGTGTGGCGGGGTACAAATGAGACCCAAACCTCGTTCTGAAATATG 378  
Db 453 ATGGGACTTTGAGATCGTGTGGCGGGGTACAAATGAGACCCAAACCTCGTTCTGAAATATG 512  
QY 379 ACGGGCCAGAAAGCGGTGTGTAACATCGCGTTCGATGAGCGCAGCTCACCTACTACG 438  
Db 513 ACGGGCCAGAAAGCGGTGTGTAACATCGCGTTCGATGAGCGCAGCTAACCCCTACTACG 572  
QY 439 CGCTGGGACCGCGAGGGGAACGCCATTGGGGGTACAGCTCGACATCGACGGGGACG 498  
Db 573 CGCTGGGACCGCGAGGGGAACGCCATTGGGGGTACAGCTCGACATCGACGGGGACG 632  
QY 499 GCGGGAGGAGATCTACTTCTCCTCAACACCAATTAATGCTTCTCGGGGGTGGCCACGTACA 558  
Db 633 GCGGGAGGAGATCTACTTCTCCTCAACACCAATTAATGCTTCTCGGGGGTGGCCACGTACA 692  
QY 559 CCGACAAGTTGTTCAAGTTCGGCAATTAACCGGTGGGAAGACATCTCGAGCGATGAGGTCA 618  
Db 693 CCGACAAGTTGTTCAAGTTCGGCAATTAACCGGTGGGAAGACATCTCGAGCGATGAGGTCA 752  
QY 619 ACGTGGCCGTGTGTGGCCAGCTCTTTGCGGAGCGCTCTGTGGCTGTGTGCGACAGAA 678  
Db 753 ACGTGGCCGTGTGTGGCCAGCTCTTTGCGGAGCGCTCTGTGGCTGTGTGCGACAGAA 812  
QY 679 AGGGCTCTGAGCCTACTCTATCTACATTGCCAATTAACCGGTGGGAAGACATCTCGAGCGATGAGGTCA 738  
Db 813 AGGGCTCTGAGCCTACTCTATCTACATTGCCAATTAACCGGTGGGAAGACATCTCGAGCGATGAGGTCA 872  
QY 739 ATGCCCTCAATGAATGGACCTTGAGCCAGTGAACCTCTCCCGGGGCAATCTCGGCGTCA 798  
Db 873 ATGCCCTCAATGAATGGACCTTGAGCCAGTGAACCTCTCCCGGGGCAATCTCGGCGTCA 932  
QY 799 GAGATGTGGCTGTGAGGCTGGGGTCAAGAAATATACAGGGGCGCAGGGCGTCAGCGTGG 858  
Db 933 GAGATGTGGCTGTGAGGCTGGGGTCAAGAAATATACAGGGGCGCAGGGCGTCAGCGTGG 992  
QY 859 GCCCATCTCTAGCAGCAGTGCCTCGGATATCTTCTCGACAAATGAGAAATGGGCCCTAACT 918  
Db 993 GCCCATCTCTAGCAGCAGTGCCTCGGATATCTTCTCGACAAATGAGAAATGGGCCCTAACT 1052  
QY 919 TCCCTTTCCAAACCGGGGGGATGGCACTTTGTGACCGTGGCGGCGAGTGTGTTGG 978  
Db 1053 TCCCTTTCCAAACCGGGGGGATGGCACTTTGTGACCGTGGCGGCGAGTGTGTTGG 1112  
QY 979 ACGACCCCAACAGCATGGCGAGTGTGCCCTGGCTGACTTCAACCGTGTGGCAAAAG 1038  
Db 1113 ACGACCCCAACAGCATGGCGAGTGTGCCCTGGCTGACTTCAACCGTGTGGCAAAAG 1172  
QY 1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCCTCTATCTGCAATGAGCACCC 1098  
Db 1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCCTCTATCTGCAATGAGCACCC 1232  
QY 1099 ATGGGAAGTCCGCTTCCGGGACATCGCTCAACCAAGTTCATCGCCCTCCCTGTCC 1158  
Db 1233 ATGGGAAGTCCGCTTCCGGGACATCGCTCAACCAAGTTCATCGCCCTCCCTGTCC 1292  
QY 1159 GCACGCTCATCACCGCGACTTTTGACAAATGACAGGAGTGGAGATCTTCTCAACAACA 1218  
Db 1293 GCACGCTCATCACCGCGACTTTTGACAAATGACAGGAGTGGAGATCTTCTCAACAACA 1352  
QY 1219 TTGCTTACCGCAGCTCTCTAGCCAAACCGCTCTTTCGCGCTCATCCGCTAGAGACGGAG 1278  
Db 1353 TTGCTTACCGCAGCTCTCTAGCCAAACCGCTCTTTCGCGCTCATCCGCTAGAGACGGAG 1412  
QY 1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCTTGGAGCCTGGAGCGCGGGGCACAG 1338  
Db 1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCTTGGAGCCTGGAGCGCGGGGCACAG 1472  
QY 1339 GGGGTGTGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTCTCCCATGGAG 1398  
Db 1473 GGGGTGTGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTGCCATGGAG 1532

QY 1399 AGTCCATGCTCAGCCGCTGTCCGTCTTCGGGGCAATCAGGGCTTCAACAACAACCTGGC 1458  
Db 1533 AGTCCATGCTCAGCCGCTGTCCGTCTTCGGGGCAATCAGGGCTTCAACAACAACCTGGC 1592  
QY 1459 TGGAGTGTGTC-ACGCAACCGGTTTGGGGCTTTTCCAGGGGAGCTAAGGTGCTGCTC 1517  
Db 1593 TGGAGTGTGTCACGCAACCGGTTTGGGGCTTTTCCAGGGGAGCTAAGGTGCTGCTC 1652  
QY 1518 TACCAACAAGAGTGGGGCCACCTGAGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577  
Db 1653 TACCAACAAGAGTGGGGCCACCTGAGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712  
QY 1578 GAGATGGAGCCCGTGGCAACATTTGGCTTGGGAAAGATGAAGCAGCAGTGTGGAGGTG 1637  
Db 1713 GAGATGGAGCCCGTGGCAACATTTGGCTTGGGAAAGATGAAGCAGCAGTGTGGAGGTG 1772  
QY 1638 ACGTGGCCAGTGGCAAGATGCTGAGCCGGAAGCTGGCCAGCGGGAGATGAACCTCAGTG 1697  
Db 1773 ACGTGGCCAGTGGCAAGATGCTGAGCCGGAAGCTGGCCAGCGGGAGATGAACCTCAGTG 1832  
QY 1698 CTGGAGATCTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCTGAGTGT 1757  
Db 1833 CTGGAGATCTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCTGAGTGT 1892  
QY 1758 GGCACAAGGATTTCTCCAGCAGGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816  
Db 1893 GGCACAAGGATTTCTCCAGCAGGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1952  
QY 1817 GTTCCCATTTGCTGTGCTCGAGACAAGCCGCTATGTCACACCTATGGAGCTACAG 1876  
Db 1953 GTTCCCATTTGCTGTGCTCGAGACAAGCCGCTATGTCACACCTATGGAGCTACAG 2012  
QY 1877 GTGCCGGAACCAACAAGTGCAGTCCGGGCTACGAGCCCCAACAGAGATGGCACAGCTG 1936  
Db 2013 GTGCCGGAACCAACAAGTGCAGTCCGGGCTACGAGCCCCAACAGAGATGGCACAGCTG 2072  
QY 1937 CGTGG 1941  
Db 2073 CGTGG 2077

Search completed: August 10, 2004, 02:06:03  
Job time : 8294 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 21:20:25 ; Search time 821 Seconds  
(without alignments)  
11264.710 Million cell updates/sec

Title: US-09-914-958B-35

Perfect score: 2177

Sequence: 1 CGGGAGGCTGAGGCCAGCC.....TCCTCTTGAAAAA 2177

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq\_29Jan04.\*

1: geneseqn1980s.\*

2: geneseqn1990s.\*

3: geneseqn2000s.\*

4: geneseqn2001as.\*

5: geneseqn2001bs.\*

6: geneseqn2002as.\*

7: geneseqn2003as.\*

8: geneseqn2003bs.\*

9: geneseqn2003cs.\*

10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177	100.0	2177	3	AA75118 cDNA enco
2	1994	91.6	2089	6	AAD29792 Human SCI
3	1759	80.8	2589	6	AAD29777 Human SCI
4	1758	80.8	2039	6	AAD29791 Human SCI
5	1758	80.8	2145	6	AAD29793 Human SCI
6	1745	80.2	1959	6	AA147918 Human cho
7	1745	80.2	2507	6	AA147917 Human cho
8	1745	80.2	2507	7	ACC50982 Human bla
9	1745	80.2	2507	7	ABX76365 Lung canc
10	1744	80.1	1947	6	AA147936 Human cho
11	1617	74.3	2413	4	AA501551 Human sec
12	1565	71.9	1849	6	AA562301 Human sequ
13	1428	65.6	2263	4	AAH23100 Osteoarth
14	1349	62.0	1574	4	AAI60925 Human pol
15	1314	60.4	1501	4	AAI59139 Human pol
16	1314	60.4	1501	8	ADB49122 Novel hum
17	1236	56.8	2178	4	AAH14000 Human CDN
18	1236	56.8	2178	7	ACC50981 Human bla
19	1236	56.8	2178	7	ABX76364 Lung canc
20	687	31.6	4794	7	ACC50983 Human bla
21	687	31.6	4794	7	ABX76366 Lung canc
22	649	29.8	883	6	ABL42156 Nucleotid
23	648	29.8	1143	3	AAZ65257 Human sec

24	648	29.8	1143	7	ADA56306 Gene enco
25	648	29.8	1143	7	ADA40142 Human sec
26	648	29.8	1143	9	ADE11646 Human sec
27	496	22.8	608	4	AAH05239 Human CDN
28	419	19.2	465	8	ACH31964 Human end
29	358	16.4	447	8	ACH42110 Human foe
c 30	191	8.8	1400	5	AA583195 DNA encod
c 31	69	3.2	319	2	AAT25375 Human gen
c 32	60	2.8	60	6	ABN43600 Human spl
c 33	57	2.6	57	6	AA147919 Human cho
c 34	41	1.9	41	6	ABL42162 Probe #1
c 35	41	1.9	41	6	ABL42161 Probe #2
c 36	40	1.8	752	4	AAH23149 Osteoarth
c 37	25	1.1	25	6	AAD29782 5' RACE p
c 38	25	1.1	33	6	ABL42159 PCR prime
c 39	24	1.1	24	6	ABL42157 PCR prime
c 40	23	1.1	23	6	AAD29784 3' RACE p
c 41	22	1.0	22	6	AAD29785 Primer #2
c 42	22	1.0	22	6	AAD29783 Primer #1
c 43	21	1.0	21	6	AA147931 Human cho
c 44	21	1.0	21	6	AAD29779 Human SCI
c 45	21	1.0	1216	6	ABQ28694 Oligonucl

## ALIGNMENTS

RESULT 1

AA75118

ID AAA75118 standard; cDNA; 2177 BP.

AC AAA75118;

XX

DT 15-JAN-2001 (first entry)

XX

DE cDNA encoding a human secretory protein.

XX

XX Human; secretory protein; HSECP; cancer; gastrointestinal disorder; inflammation; cardiovascular disorder; neurological disorder; ss.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT CDS 123..2003

FT /\*tag= a

FT /\*product= "secretory protein"

FT sig\_peptide 123..203

FT /\*tag= b

XX WO200052151-A2.

XX

XX

XX

XX PD 08-SEP-2000.

XX

XX PF 03-MAR-2000; 2000WO-US005621.

XX

XX PR 05-MAR-1999; 99US-0123117P.

XX

XX (INCY-) INCYTE PHARM INC.

XX

XX Tang YT, Lal P, Baughn MR, Yue H, Au-Young J, Lu DAM, Azimzai Y;

XX WPI; 2000-579282/54.

XX F-PSDB; AAB08864.

XX

XX Twenty two human secretory proteins for diagnosing, treating and

XX preventing cancer, inflammation, and gastrointestinal, cardiovascular and

XX neurological disorders.

XX Claim 4; Page 102-103; 107pp; English.

XX

XX The present sequence encodes a human secretory protein, designated HSECP-

XX 1. The specification also describes HSECP-2 to HSECP-22. The proteins are

XX useful for diagnosing, treating and preventing cancer, inflammation, and

XX





Db 945 GCAAAGTGGACATCGTCTATGGAACATGGATGGCCCCCAGCGCTCTATCTGCAATGA 1004  
Qy 1093 GCACCCATGGGAAGTCCGCTTCGGGACATGCGCTCAACCAAGTTCCTCCATGCCCTCCC 1152  
Db 1005 GCACCCATGGGAAGTCCGCTTCGGGACATGCGCTCAACCAAGTTCCTCCATGCCCTCCC 1064  
Qy 1153 CTGTCCGACGGTCTATCAGCCGCGACTTTGACATGACAGGAGCTGGAGATCTTCTTCA 1212  
Db 1065 CTGTCCGACGGTCTATCAGCCGCGACTTTGACATGACAGGAGCTGGAGATCTTCTTCA 1124  
Qy 1213 ACAACATTCCTTACCGGAGCTCTCAGCAACCGCTCTTCGCGCTCATCCGTAGAGAGC 1272  
Db 1125 ACAACATTCCTTACCGGAGCTCTCAGCAACCGCTCTTCGCGCTCATCCGTAGAGAGC 1184  
Qy 1273 ACGGAGACCCCTCATCGAGGAGCTCAATTCGCGGAGCGCTTTGAGCGCTTGAGGCGCGG 1332  
Db 1185 ACGGAGACCCCTCATCGAGGAGCTCAATTCGCGGAGCGCTTTGAGCGCTTGAGGCGCGG 1244  
Qy 1333 GCACAGGGGTGTGTGACCGACTTCGACGGAGAGCGGATGCTGGACCTCATCTTGTCCC 1392  
Db 1245 GCACAGGGGTGTGTGACCGACTTCGACGGAGAGCGGATGCTGGACCTCATCTTGTCCC 1304  
Qy 1393 ATGAGAGTTCATAGCTCAGCGCTGTCCGCTTCGCGGAGCAATCAGGCGTTCAACACA 1452  
Db 1305 ATGAGAGTTCATAGCTCAGCGCTGTCCGCTTCGCGGAGCAATCAGGCGTTCAACACA 1364  
Qy 1453 ACTGGCTGAGTGTGCGACACCCCGCTTTGGGCGCTTTGCCAGGGAGCTTAAGGTGG 1512  
Db 1365 ACTGGCTGAGTGTGCGACACCCCGCTTTGGGCGCTTTGCCAGGGAGCTTAAGGTGG 1424  
Qy 1513 TGCTCTACACCAAGAGTGGGCGCCACCTGAGGATCATCGCGGGCTCAGGCTACC 1572  
Db 1425 TGCTCTACACCAAGAGTGGGCGCCACCTGAGGATCATCGCGGGCTCAGGCTACC 1484  
Qy 1573 TGTGTGAGATGAGCGCGCTGCGACACTTTGGGCTGGGGAAGATGAAGCGACAGTGTGG 1632  
Db 1485 TGTGTGAGATGAGCGCGCTGCGACACTTTGGGCTGGGGAAGATGAAGCGAGTGTGG 1544  
Qy 1633 AGGTGACGTGGCGAGATGTCAGAGATGTTGAGCGGAACTGTCGACCGGGAGATGAAT 1692  
Db 1545 AGGTGACGTGGCGAGATGTCAGAGATGTTGAGCGGAACTGTCGACCGGGAGATGAAT 1604  
Qy 1693 CAGTGTGAGATTCCTTACCCCGGATGAGACACACTTCAGGACCCAGCGCCACTGG 1752  
Db 1605 CAGTGTGAGATTCCTTACCCCGGATGAGACACACTTCAGGACCCAGCGCCACTGG 1664  
Qy 1753 AGTGTGGCCAAAGATTTCTCCAGCAGGAAATGGCCATTGCGATGACACCAATGAATGA 1812  
Db 1665 AGTGTGGCCAAAGATTTCTCCAGCAGGAAATGGCCATTGCGATGACACCAATGAATGA 1724  
Qy 1813 TCAGTTCCCATTCGTGTGCGCTTCGAGACAGCCCGTATGTGTCAACACCTATGGAGCT 1872  
Db 1725 TCAGTTCCCATTCGTGTGCGCTTCGAGACAGCCCGTATGTGTCAACACCTATGGAGCT 1784  
Qy 1873 ACAGGTGCGGACCAACAGAGTGCAGTGGGGCTACGAGCCCAACAGAGATGGCACAG 1932  
Db 1785 ACAGGTGCGGACCAACAGAGTGCAGTGGGGCTACGAGCCCAACAGAGATGGCACAG 1844  
Qy 1933 CTTGCGTGGGCTGTGGAGCCCTGTGTGAAGATAGTGACACCAAGTTGGGAAGGCC 1992  
Db 1845 CTTGCGTGGGCTGTGGAGCCCTGTGTGAAGATAGTGACACCAAGTTGGGAAGGCC 1904  
Qy 1993 TTGGTCCCTGAATCACTGAATCACTGCTTGAATCAACCGCTGGAATACCTGTGTATCAG 2052  
Db 1905 TTGGTCCCTGAATCACTGAATCACTGCTTGAATCAACCGCTGGAATACCTGTGTATCAG 1964  
Qy 2053 GAACACTTACCTGGAATCACTGAGCAGATACAACTTCTATTGATTAAGCTATTAA 2112  
Db 1965 GAACACTTACCTGGAATCACTGAGCAGATACAACTTCTATTGATTAAGCTATTAA 2024  
Qy 2113 TACATTAAAGATTTGGGGGTGCTACCTTACATATAAATTCCTTCTTGAAGAAA 2172  
Db 2025 TACATTAAAGATTTGGGGGTGCTACCTTACATATAAATTCCTTCTTGAAGAAA 2084

Qy 2173 AAAAA 2177  
Db 2085 AAAAA 2089

## RESULT 3

AAD29777  
ID AAD29777 standard; cDNA; 2589 BP.

XX AAD29777;

XX AC

XX 17-MAY-2002 (first entry)

XX Human SCIM-1 protein encoding cDNA.

XX Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;  
KW extracellular matrix protein; tolerogenic protein; Grave's disease;  
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;  
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;  
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;  
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;  
KW multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 59..2020

FT /tag= a /product= "Human SCIM-1 protein"

FT sig\_peptide 59..121

FT /tag= d

FT sig\_peptide 59..115

FT /tag= b

FT mat\_peptide 116..2017

FT /tag= c

FT /product= "Human SCIM-1 mature protein #1"

FT mat\_peptide 122..2017

FT /tag= e

FT /product= "Human SCIM-1 mature protein #2"

XX WO200206478-A2.

XX 24-JAN-2002.

XX 09-JUL-2001; 2001WO-EP007888.

XX 13-JUL-2000; 2000BP-00202495.

XX (ALKU ) AKZO NOVEL NV.

XX Heus HC, Nelissen RLH, Meeuwisse CML;

XX WPI; 2002-179796/23.

XX P-PSDB; AAE18679.

XX New extracellular matrix protein useful for prevention of inflammatory diseases, more specifically in induction of T cell tolerance to the protein in patients suffering from rheumatoid arthritis.

XX Claim 4; Page 29-30; 45pp; English.

XX The invention relates to human extracellular matrix protein i.e. Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic acid molecules encoding such proteins. Sequences of the invention are useful in a screening assay for the identification of tolerogenic polypeptides. SCIM proteins are useful in therapy and for manufacture of a pharmaceutical preparation against inflammatory diseases and for induction of immunological tolerance to an autoantigen in patients suffering from autoimmune disorders, more specifically rheumatoid arthritis. The inflammatory disorders treated include glomerulonephritis, Grave's diseases, juvenile arthritis, primary glomerulonephritis, polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,









```
Db 585 ACAGAAAGGCTCTGGAGGCTACTCTATCTATCTACATTGCCAATTACGCTACCGTAAATGTGG 644
Qy 733 GCCCTGATGCCCTCATTTGAAATGACCTTGAGGCGAGTGA CTTCTCCCGGGCAATTTGG 792
Db 645 GCCCTGATGCCCTCATTTGAAATGACCTTGAGGCGAGTGA CTTCTCCCGGGCAATTTGG 704
Qy 793 CGCTCAGAGATGTGGCTCTGAGGTGGGTGAGCAAAATATACAGGGGGCGAGGCGTCA 852
Db 705 CGCTCAGAGATGTGGCTCTGAGGTGGGTGAGCAAAATATACAGGGGGCGAGGCGTCA 764
Qy 853 CGGTGGGCCCCATCTCTCAGCAGCAGTCTCGATATCTTCTGCGACAAATGAGATGGGC 912
Db 765 CGGTGGGCCCCATCTCTCAGCAGCAGTCTCGATATCTTCTGCGACAAATGAGATGGGC 824
Qy 913 CTAACCTCTCTTTTCCAAACCGGGCGAGTGGACCTTTGTGGACGCTGGGCGCAGTGTG 972
Db 825 CTAACCTCTCTTTTCCAAACCGGGCGAGTGGACCTTTGTGGACGCTGGGCGCAGTGTG 884
Qy 973 GTGTGGAGCAGCCCCACAGCAGATGGGAGGTGTGCGCTGCTGACTTCAACCGTGTG 1032
Db 885 GTGTGGAGCAGCCCCACAGCAGATGGGAGGTGTGCGCTGCTGACTTCAACCGTGTG 944
Qy 1033 GCAAAGTGGACATCTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAATGA 1092
Db 945 GCAAAGTGGACATCTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAATGA 1004
Qy 1093 GCACCATGGGAAGTCCGCTTCCGGGAGCATCGCTCACCAGTTCCTCATGCGCTGCC 1152
Db 1005 GCACCATGGGAAGTCCGCTTCCGGGAGCATCGCTCACCAGTTCCTCATGCGCTGCC 1064
Qy 1153 CTGTCCGACGCTCATACCGCGGACTTTGACAATGACAGGAGTGGAGATCTTCTTCA 1212
Db 1065 CTGTCCGACGCTCATACCGCGGACTTTGACAATGACAGGAGTGGAGATCTTCTTCA 1124
Qy 1213 ACAACATTCCTTACCGGAGTCTCTCAGCCAAACCGCTCTTCGCGCGTCACTCCGTAGAGAGC 1272
Db 1125 ACAACATTCCTTACCGGAGTCTCTCAGCCAAACCGCTCTTCGCGCGTCACTCCGTAGAGAGC 1184
Qy 1273 ACGGAGACCCCTCATTCGAGGAGTCAATCCCGGGGAGACGCTTGAGGCTTGAGGGCGGG 1332
Db 1185 ACGGAGACCCCTCATTCGAGGAGTCAATCCCGGGGAGACGCTTGAGGCTTGAGGGCGGG 1244
Qy 1333 GCACAGGGGTGTGTGACCGGACTTCGACGGAGACGGGATGCTGACCTCATCTTGTCCC 1392
Db 1245 GCACAGGGGTGTGTGACCGGACTTCGACGGAGACGGGATGCTGACCTCATCTTGTCCC 1304
Qy 1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCGGGGCAATCAGGGCTTCAACAA 1452
Db 1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCGGGGCAATCAGGGCTTCAACAA 1364
Qy 1453 ACTGGCTGGAGTGTGTCACGACCCGCTTTGGGGCTTTTGGCAGGGAGCTAAGGTG 1512
Db 1365 ACTGGCTGGAGTGTGTCACGACCCGCTTTGGGGCTTTTGGCAGGGAGCTAAGGTG 1424
Qy 1513 TGCCTTACACCAAGAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACC 1572
Db 1425 TGCCTTACACCAAGAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACC 1484
Qy 1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCTTGGGGAAGATGAAGCCAGCAGTGTG 1632
Db 1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCTTGGGGAAGATGAAGCCAGCAGTGTG 1544
Qy 1633 AGGTGACGTGGCAGATGTCAGATGTCAGCGGAACTGTCGCGGAGCTGAGGATCAACT 1692
Db 1545 AGGTGACGTGGCAGATGTCAGATGTCAGCGGAACTGTCGCGGAGCTGAGGATCAACT 1604
Qy 1693 CAGTGTGGAGATCTCTTACCCCGGGATGAGACACACTTCAGGACCCAGCCCACTGG 1752
Db 1605 CAGTGTGGAGATCTCTTACCCCGGGATGAGACACACTTCAGGACCCAGCCCACTGG 1664
Qy 1753 AGTGTGCCCAAGATTTCTCCACAGAGAAATGGCCATTTGCATGGACCAATGAATGA 1812
Db 1665 AGTGTGCCCAAGATTTCTCCACAGAGAAATGGCCATTTGCATGGACCAATGAATGA 1724
```

```
Qy 1813 TCCAGTTCCTCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
Db 1725 TCCAGTTCCTCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784
Qy 1873 ACAGTGTCCGACCAACAAGAAAGTGCAGTCCGGGGCTACGAGCCCAACGAGGATGCGACAG 1932
Db 1785 ACAGTGTCCGACCAACAAGAAAGTGCAGTCCGGGGCTACGAGCCCAACGAGGATGCGACAG 1844
Qy 1933 CCTGCGTGG 1941
Db 1845 CCTGCGTGG 1853

RESULT 6
AAL47918
ID AAL47918 standard; cDNA; 1959 BP.
XX
AC AAL47918;
XX
DT 26-SEP-2002 (first entry)
XX
DE Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:2.
XX
KW Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW extracellular matrix protein; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1959
FT /tag= a
FT /product= "CEP-68"
FT /partial
FT /note= "no stop codon"
FT sig_peptide 1..57
FT mat_peptide 58..1959
FT /tag= c
XX
PN WO200253709-A1.
XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-BP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
XX WPI; 2002-528856/56.
DR P-PSDB; AA018264.
XX
XX New nucleic acid encoding an extracellular matrix protein, useful, for
example, as a marker for detecting or isolating chondrocytes.
PS Claim 2; Page 92; 107pp; German.
XX
XX The present invention relates to the protein and coding sequences of
novel human extracellular matrix proteins, more specifically chondrocyte
specific CEP-68 proteins. The sequences can be used in the identification
of cell differentiation stages in cells which develop from mesenchymal
stem cells, particularly chondrocytes and to isolate chondrocytes. The
present sequence is a coding sequence of the CEP-68 protein
XX
XX Sequence 1959 BP; 394 A; 590 C; 596 G; 379 T; 0 U; 0 Other;
Query Match 80.2%; Score 1745; DB 6; Length 1959;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy	147	ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCGCCCATACTGAGGGTCC	206
Db	1	ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCGCCCATACTGAGGGTCC	60
Qy	207	CAGCGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCTCTGACATATGAC	266
Db	61	CAGCGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCTCTGACATATGAC	120
Qy	267	AGTAATCCCAACCCAGCTCAACTATGGTGTGGCACTTACTGATGTGGACCATGATGGGGAC	326
Db	121	AGTAATCCCAACCCAGCTCAACTATGGTGTGGCACTTACTGATGTGGACCATGATGGGGAC	180
Qy	327	TTTGAGATCGTGTGGGGGGTACAATGGACCCAACTGTGTTCTGAAGTATGACCGGGCC	386
Db	181	TTTGAGATCGTGTGGGGGGTACAATGGACCCAACTGTGTTCTGAAGTATGACCGGGCC	240
Qy	387	CAGAACGGCTGTGAACATCGCGTGGATGAGCGCAGCTCACCTACTAGCGCTCGGG	446
Db	241	CAGAACGGCTGTGAACATCGCGTGGATGAGCGCAGCTCACCTACTAGCGCTCGGG	300
Qy	447	GACCGGACGGGAAACGCCATTTGGGCTCACAGCCTGCGACATCGACGGGGACGGCGGGAG	506
Db	301	GACCGGACGGGAAACGCCATCGGGGTACAGCCTGCGACATCGACGGGGACGGCGGGAG	360
Qy	507	GAGATCTACTTCTCAACAAATATAGCCTTCTCGGGGTGGCCACGTACCGACAAG	566
Db	361	GAGATCTACTTCTCAACAAATATAGCCTTCTCGGGGTGGCCACGTACCGACAAG	420
Qy	567	TTGTTCAGTTCCGCAATAACGGTGGGAGACATCTGAGCGATGAGGTCAAGTGGCC	626
Db	421	TTGTTCAGTTCCGCAATAACGGTGGGAGACATCTGAGCGATGAGGTCAAGTGGCC	480
Qy	627	CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGCTGTGTGFGACAGAAAGGGCTCT	686
Db	481	CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGCTGTGTGFGACAGAAAGGGCTCT	540
Qy	687	GGACGTACTCTATCTCAATTCGCCAATTACGCCCTACGGTAAATGTGGGCCCTGATGCCCTC	746
Db	541	GGACGTACTCTATCTCAATTCGCCAATTACGCCCTACGGTAAATGTGGGCCCTGATGCCCTC	600
Qy	747	ATTGAATGACCTGAGGCCAGTGACTCTCCGGGGCATTCGCGCGCTCAGAGATGTG	806
Db	601	ATTGAATGACCTGAGGCCAGTGACTCTCCGGGGCATTCGCGCGCTCAGAGATGTG	660
Qy	807	GCTGCTGAGGCTGGGGTCAGCAAAATATACAGGGGGCCGAGCGCTACGCTGGGGCCCCATC	866
Db	661	GCTGCTGAGGCTGGGGTCAGCAAAATATACAGGGGGCCGAGCGCTACGCTGGGGCCCCATC	720
Qy	867	CTCAGCAGCATGCGCTCGGATATCTTCTGGACAAATGAGAAATGGGCCCTAACTCTCTTTTC	926
Db	721	CTCAGCAGCATGCGCTCGGATATCTTCTGGACAAATGAGAAATGGGCCCTAACTCTCTTTTC	780
Qy	927	CACAAACGGGGCATGACCTTTGTGGACGCTCGGCCAGTCTGCTGTGTGGACACCC	986
Db	781	CACAAACGGGGCATGACCTTTGTGGACGCTCGGCCAGTCTGCTGTGTGGACACCC	840
Qy	987	CACCAGCATGGGGAGGTGTGCGCCTCGCTGACTTTCAACCGTGATGCCAAAGTGGACATC	1046
Db	841	CACCAGCATGGGGAGGTGTGCGCCTCGCTGACTTTCAACCGTGATGCCAAAGTGGACATC	900
Qy	1047	GTCATTGGCAACTGGAATGGCCCCACCGCTCTATCTGCAAAATGACACCCCATGGGAAG	1106
Db	901	GTCATTGGCAACTGGAATGGCCCCACCGCTCTATCTGCAAAATGACACCCCATGGGAAG	960
Qy	1107	GTCGCGCTTCGGGACATGCGCTCACCAAGTTCTTCATGCGCTCCCTCTGTCGCGACGGTCTC	1166
Db	961	GTCGCGCTTCGGGACATGCGCTCACCAAGTTCTTCATGCGCTCCCTCTGTCGCGACGGTCTC	1020
Qy	1167	ATCACCGCCGACTTTGACAATGACACGAGCTGGAGATCTTCTTCAACAACATTCCTAC	1226
Db	1021	ATCACCGCCGACTTTGACAATGACACGAGCTGGAGATCTTCTTCAACAACATTCCTAC	1080
Qy	1227	CGCAGCTCCTCAGCAACCGCCTCTTCGGGTCATCTCGGTAGAGACACGGAGACCCCTCTC	1286

Db	1081	CGCAGCTCCTCAGCAACCGCGCTTTTCGGGGTTCATCGTAGAGACACGGAGACCCCTC	1140
QY	1287	ATCGAGAGAGTCAATCCCGGCGAGCGCTTCGAGCGTCGAGGCGCGGGCCAGAGGGGGTGTG	1346
Db	1141	ATCGAGAGAGTCAATCCCGGCGAGCGCTTCGAGCGTCGAGGCGCGGGCCAGAGGGGGTGTG	1200
QY	1347	GTGACCGAGTTCGACCGAGACGGGATGCTGGACCTCATCTTTGTGCCATGGAGAGTCCATG	1406
Db	1201	GTGACCGAGTTCGACCGAGACGGGATGCTGGACCTCATCTTTGTGCCATGGAGAGTCCATG	1260
QY	1407	GCTCAGCCGCTGTCCGTTCCTCCGGGCAATCAAGGCTTCAACAACAACACTGGCTGCGAGTG	1466
Db	1261	GCTCAGCCGCTGTCCGTTCCTCCGGGCAATCAAGGCTTCAACAACAACACTGGCTGCGAGTG	1320
QY	1467	GTGCCACGCAACCCGGTTTGGGGCTTTGCCAGGGGAGCTAAGTTCGTCTACACCAAG	1526
Db	1321	GTGCCACGCAACCCGGTTTGGGGCTTTGCCAGGGGAGCTAAGTTCGTCTACACCAAG	1380
QY	1527	AAGAGTGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG	1586
Db	1381	AAGAGTGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG	1440
QY	1587	CCGTGGCAACATTTGGCCTTGGGGAAGGATGAAGCAGCAGTGTGAGAGTGAAGTGGCCA	1646
Db	1441	CCGTGGCAACATTTGGCCTTGGGGAAGGATGAAGCAGCAGTGTGAGAGTGAAGTGGCCA	1500
QY	1647	GATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTCTGGAGATC	1706
Db	1501	GATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTCTGGAGATC	1560
QY	1707	CTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGA	1766
Db	1561	CTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGA	1620
QY	1767	TTCCTCCAGCAGGAATAATGGCCATTCATGGGACACCAATGAATGCATCCAGTTCCTCATTC	1826
Db	1621	TTCCTCCAGCAGGAATAATGGCCATTCATGGGACACCAATGAATGCATCCAGTTCCTCATTC	1680
QY	1827	GTGTGCCCTCGACACAGCCCGTATGTGTCAACACCTATGGAGCTACAGTCCCGGACC	1886
Db	1681	GTGTGCCCTCGACACAGCCCGTATGTGTCAACACCTATGGAGCTACAGTCCCGGACC	1740
QY	1887	AACAAGAGTGCAGTCCGGGCTACGAGCCCAACGAGGATGGCACAGCTCGCTGGG	1942
Db	1741	AACAAGAGTGCAGTCCGGGCTACGAGCCCAACGAGGATGGCACAGCTCGCTGGG	1796
RESULT 7			
AAL47917			
ID AAL47917 standard; cDNA; 2507 BP.			
AC	AAL47917;		
XX			
XX	26-SEP-2002 (first entry)		
XX	Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:1.		
DE			
XX	Human; chondrocyte specific protein; CEP-68; cartilage; marker;		
KW	extracellular matrix protein; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
FE	Key	Location/Qualifiers	
FT	CDS	1..1962	
FT		/*tag= a	
FT		/product= "CEP-68"	
FT	sig_peptide	1..57	
FT		/*tag= b	
FT	mat_peptide	58..1959	
FT		/*tag= c	
XX			
XX	W0200253709-A1.		

















QY 1459 TCGAGTGTGTCACGACCCCGGTTTGGGCGCTTTGCCAGGGAGCTAAGGTGCTGCT 1518  
DB 1758 TCGAGTGTGTCACGACCCCGGTTTGGGCGCTTTGCCAGGGAGCTAAGGTGCTGCT 1817  
QY 1519 ACACCAAGAAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGTACCTGTGTG 1578  
DB 1818 ACACCAAGAAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGTACCTGTGTG 1877  
QY 1579 AGATGAGCCCGTGGGACACTTTGGCTGGGAAGGATGAAGCCAGTGTGGAGTGA 1638  
DB 1878 AGATGAGCCCGTGGGACACTTTGGCTGGGAAGGATGAAGCCAGTGTGGAGTGA 1937  
QY 1639 CTGGCCAGATGCGAAGTGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGC 1698  
DB 1938 CTGGCCAGATGCGAAGTGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGC 1997  
QY 1699 TGGAGATCCTTACCCCGGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTG 1758  
DB 1998 TGGAGATCCTTACCCCGGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTG 2057  
QY 1759 GCCAAGGATTCCTCCAGCAGGAAATGGCCATTCGATGACACCAATGAATGCCAGT 1818  
DB 2058 GCCAAGGATTCCTCCAGCAGGAAATGGCCATTCGATGACACCAATGAATGCCAGT 2117  
QY 1819 TCCCATTCGTGTGCGCTCGAGACAGCCCGTATGTGTCAACACCTATGGAAGTACAGGT 1878  
DB 2118 TCCCATTCGTGTGCGCTCGAGACAGCCCGTATGTGTCAACACCTATGGAAGTACAGGT 2177  
QY 1879 GCCGACCAACAAGAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGCCACAGCTGCG 1938  
DB 2178 GCCGACCAACAAGAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGCCACAGCTGCG 2237  
QY 1939 TGG 1941  
DB 2238 TGG 2240

## RESULT 12

AAS62301

ID AAS62301 standard; cDNA; 1849 BP.

AC AAS62301;

DT 14-FEB-2002 (first entry)

DE cDNA sequence #88 encoding novel human secreted protein.

XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
KW immune deficiency disorder; blood disorder; inflammatory disorder;  
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
KW immunosuppressive; antirheumatic; ss.

OS Homo sapiens.

XX WO200177291-A2.

XX 18-OCT-2001.

XX 29-MAR-2001; 2001WO-US010485.

XX 06-APR-2000; 2000US-0195604P.

XX (GEMY ) GENETICS INST INC.

XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;

PI Gulukota K, Graham JR;

XX WPI; 2002-010900/01.

XX New polynucleotides encoding secreted proteins useful for treating e.g.

PT asthma, HIV and Crohn's disease.

XX

PS Claim 1; Page 122; 391pp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
CC which encode human secreted proteins. The cDNA sequences have been  
CC derived from a variety of human tissues. The invention also provides a  
CC method for producing proteins from these polynucleotide sequences. The  
CC proteins are useful for identifying compounds that modulate their  
CC activity and production, and the cell is also useful for identifying  
CC compounds that modulate expression of the polynucleotide sequences  
CC encoding the secreted proteins. The sequences of the invention are useful  
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
CC immune deficiency disorders (e.g. severe combined immunodeficiency  
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
CC the invention are also useful in gene therapy. AAS62214-AAS62838  
CC represent the cDNA sequences of the invention that encode for novel human  
CC secreted proteins  
XX

SQ Sequence 1849 BP; 395 A; 535 C; 544 G; 372 T; 0 U; 3 Other;

Query Match 71.9%; Score 1565; DB 6; Length 1849;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1665; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 239 CAACTCAGTTCGTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGC 238

DB 1 CAACTCAGTTCGTCTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGC 60

QY 299 AGTTACTGATGTGGACATGATGGGACTTTGAGATCGTGTGGGGGTACAATGGACC 358

DB 61 AGTTACTGATGTGGACATGATGGGACTTTGAGATCGTGTGGGGGTACAATGGACC 120

QY 359 CAACTGTTTCTGAAGTATGACCGGGCCAGAAAGCGGTGGTGAACATCCGGTTCATGA 418

DB 121 CAACTGTTTCTGAAGTATGACCGGGCCAGAAAGCGGTGGTGAACATCCGGTTCATGA 180

QY 419 GCGCAGCTCACCTCTACGCGCTCGGGACCGGAGGAGAACCCATTGGGTTCACAGC 478

DB 181 GCGCAGCTCACCTCTACGCGCTCGGGACCGGAGGAGAACCCATTGGGTTCACAGC 240

QY 479 CTGCGACATCGACGGGACCGGAGGAGATCTACTTCCTCAACACCAATAATGCCCTT 538

DB 241 CTGCGACATCGACGGGACCGGAGGAGATCTACTTCCTCAACACCAATAATGCCCTT 300

QY 539 CTCGGGGTGGCCAGTACACCGACAAGTTGTTCAAGTTCGCAATAACCGTGGGAAGA 598

DB 301 CTCGGGGTGGCCAGTACACCGACAAGTTGTTCAAGTTCGCAATAACCGTGGGAAGA 360

QY 599 CATCTGAGCGATGAGGTCAACGTGGCCCGTGGTGGCCAGCTCTTTGCCGAGCGCTC 658

DB 361 CATCTGAGCGATGAGGTCAACGTGGCCCGTGGTGGCCAGCTCTTTGCCGAGCGCTC 420

QY 659 TGTGCCCTGTGTGGACAGAAAGGCTCTGGAGCTACTCTATCTACATTCGCAATTCACGC 718

DB 421 TGTGCCCTGTGTGGACAGAAAGGCTCTGGAGCTACTCTATCTACATTCGCAATTCACGC 480

QY 719 CTACCGTAAATGTGGCCCTGATGCCCTCATTTGAATGGACCTGAGGCGAGTCTCTC 778

DB 481 CTACCGTAAATGTGGCCCTGATGCCCTCATTTGAATGGACCTGAGGCGAGTCTCTC 540

QY 779 CCGGGGCATTTCTGGCGCTCAGAGATGTGGCTCTGAGGCTGGGGTTCAGCAATATACAGG 838

DB 541 CCGGGGCATTTCTGGCGCTCAGAGATGTGGCTCTGAGGCTGGGGTTCAGCAATATACAGG 600

QY 839 GGGCGCAGCGTTCAGCGTGGGCCCCATCTCTAGAGAGAGTCTCGGATATCTTCGCGA 898

DB 601 GGGCGCAGCGTTCAGCGTGGGCCCCATCTCTAGAGAGAGTCTCGGATATCTTCGCGA 660

QY 899 CAACTCAGAAATGGCCCTAACTTCTTTTCCACAACCGGGCGATGGCACTTTGTGACGC 958

DB 661 CAACTCAGAAATGGCCCTAACTTCTTTTCCACAACCGGGCGATGGCACTTTGTGACGC 720



Db 513 ACCGGGCCAGAAAGCGCTGGTGAACATCGCGTTCGATGAGCGCAGTAAACCCCTACTACG 572

Qy 439 CGCTGGGACCGGAGGGAAGCCATTGGGGTCAACAGCTTCGACATCGACGGGAGC 498

Db 573 CGCTGGGACCGGAGGGAAGCCATCGGGGTCAACAGCTTCGACATCGACGGGAGC 632

Qy 499 GCGGGAGGAGATCTACTTCTCAACACCAATAAATGCCCTTCGCGGGGTGGCCAGCTACA 558

Db 633 GCGGGAGGAGATCTACTTCTCAACACCAATAAATGCCCTTCGCGGGGTGGCCAGCTACA 692

Qy 559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCTGACGATGAGGTCA 618

Db 693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCTGACGATGAGGTCA 752

Qy 619 AGTGGCCCGTGGTGGCCAGCCTCTTTGCCGAGCGCTCTGTGGCTGTGTGGACAGAA 678

Db 753 AGTGGCCCGTGGTGGCCAGCCTCTTTGCCGAGCGCTCTGTGGCTGTGTGGACAGAA 812

Qy 679 AGGGCTCTGGACGCTACTCTATCTACATTCGCAATTCGACCTACGGTAATGTGGGCCCTG 738

Db 813 AGGGCTCTGGACGCTACTCTATCTACATTCGCAATTCGACCTACGGTAATGTGGGCCCTG 872

Qy 739 ATGCCCTCATTAAGATGACCTGAGCGCAGTGAACCTCTCCGGGGCAATCTTGGGCTCA 798

Db 873 ATGCCCTCATTAAGATGACCTGAGCGCAGTGAACCTCTCCGGGGCAATCTTGGGCTCA 932

Qy 799 GAGATGTGGCTCTGAGGCTGGGTGAGCAATATACAGGGGGCCGAGCGCTCAGCGTG 858

Db 933 GAGATGTGGCTCTGAGGCTGGGTGAGCAATATACAGGGGGCCGAGCGCTCAGCGTG 992

Qy 859 GCCCATCTCAGCAGCAGTGCTCGGATATCTCTCGCACAATGAGAAATGGCCCTAACT 918

Db 993 GCCCATCTCAGCAGCAGTGCTCGGATATCTCTCGCACAATGAGAAATGGCCCTAACT 1052

Qy 919 TCCTTTTCCAAACCGGGCGATGGCACTTTGTGGAGCTCCGCCAGTGTGTTGG 978

Db 1053 TCCTTTTCCAAACCGGGCGATGGCACTTTGTGGAGCTCCGCCAGTGTGTTGG 1112

Qy 979 AGCAACCCACACAGATGGGAGGTGCGCCCTGGCTGACTTCAACCGTGTGCAAG 1038

Db 1113 AGCAACCCACACAGATGGGAGGTGCGCCCTGGCTGACTTCAACCGTGTGCAAG 1172

Qy 1039 TGGACATCGTCTATGGCACTGGAATGGCCCGCCAGCTCTATCTGCAAAATGAGCACCC 1098

Db 1173 TGGACATCGTCTATGGCACTGGAATGGCCCGCCAGCTCTATCTGCAAAATGAGCACCC 1232

Qy 1099 ATGGAAAGTTCGCTTCGGGACATCGCCTCAACCAAGTTCTCAATGCCCTCCCTGTCC 1158

Db 1233 ATGGAAAGTTCGCTTCGGGACATCGCCTCAACCAAGTTCTCAATGCCCTCCCTGTCC 1292

Qy 1159 GCACGCTCATCACCCGCACTTTGACAATGACCGAGGTGGAGATCTCTTCAACAACA 1218

Db 1293 GCACGCTCATCACCCGCACTTTGACAATGACCGAGGTGGAGATCTCTTCAACAACA 1352

Qy 1219 TTGCTTACCGAGCTCTCAGCAACCGCTCTTCGCGCTCATCGTAGAGACGCGAG 1278

Db 1353 TTGCTTACCGAGCTCTCAGCAACCGCTCTTCGCGCTCATCGTAGAGACGCGAG 1412

Qy 1279 ACCCTCTCATCAGAGCTCAATCCGGGAGCGCTTGAGCCTGAGGCGCGGGACAG 1338

Db 1413 ACCCTCTCATCAGAGCTCAATCCGGGAGCGCTTGAGCCTGAGGCGCGGGACAG 1472

Qy 1339 GGGGTGTGTGACCGACTTCGACGAGAGCGGATGCTGACCTCATCTTGTCCCATGGAG 1398

Db 1473 GGGGTGTGTGACCGACTTCGACGAGAGCGGATGCTGACCTCATCTTGTCCCATGGAG 1532

Qy 1399 AGTCAATGCTCAGCGGTGTCCTCTTCGCGGGCAATCAGGGCTTCAACAACAATGCGC 1458

Db 1533 AGTCAATGCTCAGCGGTGTCCTCTTCGCGGGCAATCAGGGCTTCAACAACAATGCGC 1592

Qy 1459 TCGGAGTGTGTC-ACGACCCCGTTTGGGCTTTTGGCGGCTTTTGGCGGAGCTAAAGTCTGTCT 1517

Db 1593 TCGGAGTGTGTC-ACGACCCCGTTTGGGCTTTTGGGCTTTTGGCGGAGCTAAAGTCTGTCT 1652

Qy 1518 TACACCAAGAGAGTGGGSCCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGT 1577

Db 1653 TACACCAAGAGAGTGGGSCCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGT 1712

Qy 1578 GAGATGGAGCCGCTGGCACACATTTGGCCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637

Db 1713 GAGATGGAGCCGCTGGCACACATTTGGCCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772

Qy 1638 ACGTGGCCAGATGGCAAGATGTTGAGCGCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTG 1697

Db 1773 ACGTGGCCAGATGGCAAGATGTTGAGCGCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTG 1832

Qy 1698 CTGGAGATCTCTACCCCGGATGAGGACACACTTTCAGGACCCAGCCACTGGAGTGT 1757

Db 1833 CTGGAGATCTCTACCCCGGATGAGGACACACTTTCAGGACCCAGCCACTGGAGTGT 1892

Qy 1758 GGCCAAAGATTTCTCCAGCAGGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816

Db 1893 GGCCAAAGATTTCTCCAGCAGGAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952

Qy 1817 GTTCCCATTCGTGTGGCTTCGAGCAAGCCCGTATGTGTCAACACTATGGAAGCTACAG 1876

Db 1953 GTTCCCATTCGTGTGGCTTCGAGCAAGCCCGTATGTGTCAACACTATGGAAGCTACAG 2012

Qy 1877 GTCCCGGACCAACAAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCTG 1936

Db 2013 GTCCCGGACCAACAAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCTG 2072

Qy 1937 CGTGG 1941

Db 2073 CGTGG 2077

RESULT 14

AAI60925

ID AAI60925 standard; cDNA; 1574 BP.

XX AAI60925;

AC AAI60925;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4914.

DE Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX Homo sapiens.

OS WC200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

PI

PI	Zhou P, Goodrich R, Drmanac RT;	
XX	WPI; 2001-442253/47.	
DR	P-PSDB; AAM41769.	
XX	Novel nucleic acids and polypeptides, useful for treating disorders such	
PT	as central nervous system injuries.	
XX	Claim 1; SEQ ID NO 4914; 10078pp; English.	
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and the	
CC	encoded polypeptides (AAM38642-AAM42213) with nootropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	C.N.S disorders. Note: The sequence data for this patent did not form	
CC	part of the printed specification	
XX	Sequence 1574 BP; 358 A; 445 C; 455 G; 316 T; 0 U; 0 Other;	
SQ	Query Match 62.0%; Score 1349; DB 4; Length 1574;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	593 GGAAGACATCCTGAGCGATGAGTCAACGTGGCCGCTGGTGGCCAGCCCTTTTGGCGG 652	
DB	1 GGAAGACATCCTGAGCGATGAGTCAACGTGGCCGCTGGTGGCCAGCCCTTTTGGCGG 60	
QY	653 ACCTCTGTGGCTGTGGACGAAAGGGCTGTGGACGCTACTCTATCTACATTTGCCAA 712	
DB	61 ACCTCTGTGGCTGTGGACGAAAGGGCTGTGGACGCTACTCTATCTACATTTGCCAA 120	
QY	713 TTACGCTACGTAATGTGGCCCTGATGCCCTCATTGAATGACCCCTCAGCGCCAGTGA 772	
DB	121 TTACGCTACGTAATGTGGCCCTGATGCCCTCATTGAATGACCCCTCAGCGCCAGTGA 180	
QY	773 CCTCTCCGGGGCATTTCCGGCTCAGAGATGGCTGTGGCTGGGCTGGGCTCAGCAATA 832	
DB	181 CCTCTCCGGGGCATTTCCGGCTCAGAGATGGCTGTGGCTGGGCTGGGCTCAGCAATA 240	
QY	833 TACAGGGGGCGGCGCTCAGCGTGGGCCCGCATCTCTCAGCAGCATGTCCTCGGATATCTT 892	
DB	241 TACAGGGGGCGGCGCTCAGCGTGGGCCCGCATCTCTCAGCAGCATGTCCTCGGATATCTT 300	
QY	893 CTGCGCAATGAGATGGGCTAACTTCCTTTTCCACACCGGGCGCATGSCACTTTGT 952	
DB	301 CTGCGCAATGAGATGGGCTAACTTCCTTTTCCACACCGGGCGCATGSCACTTTGT 360	
QY	953 GGACGCTGCGGCCAGTCTGTGTGGACGACCCGCCACAGCATGGCGAGGTGTGCGCCT 1012	
DB	361 GGACGCTGCGGCCAGTCTGTGTGGACGACCCGCCACAGCATGGCGAGGTGTGCGCCT 420	
QY	1013 GGCTGATTTCAACCGTGAATGGCAAGTGGACATCGTCTATGGCAATGGCGCCGCCA 1072	
DB	421 GGCTGATTTCAACCGTGAATGGCAAGTGGACATCGTCTATGGCAATGGCGCCGCCA 480	
QY	1073 CCGGCTCTATCTCAATGACACCCATGGGAAGTCCGCTTCGGGACATCGCTCACC 1132	
DB	481 CCGGCTCTATCTCAATGACACCCATGGGAAGTCCGCTTCGGGACATCGCTCACC 540	
QY	1133 CAAGTTCTCCATGCCCTCCCTGTCGCGACGCGTCATCACCGCCGACTTTGACAAATGACCA 1192	
DB	541 CAAGTTCTCCATGCCCTCCCTGTCGCGACGCGTCATCACCGCCGACTTTGACAAATGACCA 600	
QY	1193 GGAGCTGGAGATCTTCTTCAACAAATTGCTACCGAGTCTCTCAGCCACCGCCCTCTT 1252	

Db	601 GGAGCTGGAGATCTTCTTCAACAACTTGCTTACCGAGCTCTCTCAACCGCCCTCTT 660	
QY	1253 CCGCGTCATCCGTAGAGACACGAGACCCCTCATCGAGAGCTCAATCCCGCGACGC 1312	
Db	661 CCGCGTCATCCGTAGAGACACGAGACCCCTCATCGAGAGCTCAATCCCGCGACGC 720	
QY	1313 CTTGGAGCCTTAGGGCCGGGGCACAGGGGTGTGGTGAACCGACTTTCGACGAGACGGAT 1372	
Db	721 CTTGGAGCCTTAGGGCCGGGGCACAGGGGTGTGGTGAACCGACTTTCGACGAGACGGAT 780	
QY	1373 GCTGGACCTCATCTTGTCCATGAGAGTCCATGGGCTCAGCCGCTGCTCCCTTCCGGGG 1432	
Db	781 GCTGGACCTCATCTTGTCCATGAGAGTCCATGGGCTCAGCCGCTGCTCCCTTCCGGGG 840	
QY	1433 CAATCAGGGCTTCAACAACTGGCTGGAGTGGTGGCCACGACCCGGTTTGGGGCCTT 1492	
Db	841 CAATCAGGGCTTCAACAACTGGCTGGAGTGGTGGCCACGACCCGGTTTGGGGCCTT 900	
QY	1493 TGCACGGGAGCTAAGGTGTGTCTTACCAAGAGAGTGGGGCCACCTGAGGATCAT 1552	
Db	901 TGCACGGGAGCTAAGGTGTGTCTTACCAAGAGAGTGGGGCCACCTGAGGATCAT 960	
QY	1553 CGACGGGGCTCAGGCTACCTGTGTGAGTGGAGCCGCTGGCACATTTTGGCCTGGGAA 1612	
Db	961 CGACGGGGCTCAGGCTACCTGTGTGAGTGGAGCCGCTGGCACATTTTGGCCTGGGAA 1020	
QY	1613 GGATGAAGCCAGCAGTGTGGAGTGCAGTGGCCAGATGGCAAGATGGTGGCCGGAACGT 1672	
Db	1021 GGATGAAGCCAGCAGTGTGGAGTGCAGTGGCCAGATGGCAAGATGGTGGCCGGAACGT 1080	
QY	1673 GGCACGGGGAGATGAATCAGTGTGTGGAGATCTCTACCCCGGGATGAGGACACACT 1732	
Db	1081 GGCACGGGGAGATGAATCAGTGTGTGGAGATCTCTACCCCGGGATGAGGACACACT 1140	
QY	1733 TCAGGACCCAGCCCCACTGGAGTGGCCAGAGATTTCCACGAGCAAAATGGCCATTG 1792	
Db	1141 TCAGGACCCAGCCCCACTGGAGTGGCCAGAGATTTCCACGAGCAAAATGGCCATTG 1200	
QY	1793 CATGGACACCAATGAATGATCCAGTTCCTTCCATTCGTGTGCCCTCGACACAAGCCCGTATG 1852	
Db	1201 CATGGACACCAATGAATGATCCAGTTCCTTCCATTCGTGTGCCCTCGACACAAGCCCGTATG 1260	
QY	1853 TGTCAACACCTATGGAAGCTACAGGTGCCGACCAACAAGAGTGCAGTGGGGCTACGA 1912	
Db	1261 TGTCAACACCTATGGAAGCTACAGGTGCCGACCAACAAGAGTGCAGTGGGGCTACGA 1320	
QY	1913 GCCCAAGAGGATGGCACAGCCTGCGTGG 1941	
Db	1321 GCCCAAGAGGATGGCACAGCCTGCGTGG 1349	
RESULT 15		
AAI59139		
ID	AAI59139 standard; cDNA; 1501 BP.	
XX	AAI59139;	
AC	AC	
XX	22-OCT-2001 (first entry)	
DT		
XX	Human polynucleotide SEQ ID NO 1342.	
DE		
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	leukaemia; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200153312-A1.	
XX		

PD 26-JUL-2001.  
XX 26-DEC-2000; 2000WO-US034263.  
XX 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
P-PSDB; AAM39983.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
XX Claim 1; SEQ ID NO 1342; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: the sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 1501 BP; 319 A; 436 C; 441 G; 305 T; 0 U; 0 Other;  
Query Match 60.4%; Score 1314; DB 4; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 628 GTGGTGTGGCCAGCCTCTTTGGCGACGCTCTGTGGCCCTGTGGACAGAAAGGGCTCTG 687  
DB 36 GTGGTGTGGCCAGCCTCTTTGGCGACGCTCTGTGGCCCTGTGGACAGAAAGGGCTCTG 95  
QY 688 GACGCTACTCTATCTACATTTGCAATTAAGCTACGCTAATGTGGCCCTGATGCCCTCA 747  
DB 96 GACGCTACTCTATCTACATTTGCAATTAAGCTACGCTAATGTGGCCCTGATGCCCTCA 155  
QY 748 TTGAATGACCCCTGAGCCAGTACCTCTCCCGGGGATCTGGCGCTCAGAGATGTGG 807  
DB 156 TTGAATGACCCCTGAGCCAGTACCTCTCCCGGGGATCTGGCGCTCAGAGATGTGG 215  
QY 808 CTGCTGAGGCTGGGCTCAGCAATATACAGGGGGCGAGGCGCTCAGCGTGGGCCCATCC 867  
DB 216 CTGCTGAGGCTGGGCTCAGCAATATACAGGGGGCGAGGCGCTCAGCGTGGGCCCATCC 275  
QY 868 TCAGCAGCAGTGCCTCGGATATCTTCTGCGCAATAGAGATGGGCGCTAATCTCTTTCC 927  
DB 276 TCAGCAGCAGTGCCTCGGATATCTTCTGCGCAATAGAGATGGGCGCTAATCTCTTTCC 335  
QY 928 ACAACCGGGCGATGGCACTTTGTGGACGCTCGCGCCAGTGTGGTGGACGACCCGCC 987  
DB 336 ACAACCGGGCGATGGCACTTTGTGGACGCTCGCGCCAGTGTGGTGGACGACCCGCC 395

QY 988 ACCAGCATGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTGTATGCAAGTGGACATCG 1047  
DB 396 ACCAGCATGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTGTATGCAAGTGGACATCG 455  
QY 1048 TCTATGGCAACTGGAATGGCCCGCCACCGCTCTATCTGCAAAATGAGCACCCATGGGAAGG 1107  
DB 456 TCTATGGCAACTGGAATGGCCCGCCACCGCTCTATCTGCAAAATGAGCACCCATGGGAAGG 515  
QY 1108 TCGCTTCCGGGACATCGGCTCAACCAAGTTCTCATGCGCTTCCCTGTCGCGACGGTCA 1167  
DB 516 TCGCTTCCGGGACATCGGCTCAACCAAGTTCTCATGCGCTTCCCTGTCGCGACGGTCA 575  
QY 1168 TCACCGCCGACTTTCACATGACACGAGGAGCTGAGATCTTCTCAACAACATTTGCCCTACC 1227  
DB 576 TCACCGCCGACTTTCACATGACACGAGGAGCTGAGATCTTCTCAACAACATTTGCCCTACC 635  
QY 1228 GCAGCTCTCTCAGCCAAACCGCTCTTCCGCGCTCATCGTAGAGAGACGAGACCCGCTCA 1287  
DB 636 GCAGCTCTCTCAGCCAAACCGCTCTTCCGCGCTCATCGTAGAGAGACGAGACCCGCTCA 695  
QY 1288 TCGAGGAGCTCAATCCCGCGAGCGCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 1347  
DB 696 TCGAGGAGCTCAATCCCGCGAGCGCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 755  
QY 1348 TGACCGACTTCGACGAGACGCGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 1407  
DB 756 TGACCGACTTCGACGAGACGCGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 815  
QY 1408 CTCAGCGCTGTCCGCTTTCCGGGGCAATCAGGGCTTCAACAACAACCTGGCTGCGAGTGG 1467  
DB 816 CTCAGCGCTGTCCGCTTTCCGGGGCAATCAGGGCTTCAACAACAACCTGGCTGCGAGTGG 875  
QY 1468 TCCACGACCCCGCTTTGGGGCTTTGCCAGGGGAGCTAAGTCTGCTCTACACCAAGA 1527  
DB 876 TCCACGACCCCGCTTTGGGGCTTTGCCAGGGGAGCTAAGTCTGCTCTACACCAAGA 935  
QY 1528 AGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587  
DB 936 AGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995  
QY 1588 CCGTGGCACACTTTGGCCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGCCAG 1647  
DB 996 CCGTGGCACACTTTGGCCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGCCAG 1055  
QY 1648 ATGGCAAGATGCTGAGCCGGAACGCTGGCCAGCGGGGAGATGAATCAGTCTGGAGATCC 1707  
DB 1056 ATGGCAAGATGCTGAGCCGGAACGCTGGCCAGCGGGGAGATGAATCAGTCTGGAGATCC 1115  
QY 1708 TCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAGGAT 1767  
DB 1116 TCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAGGAT 1175  
QY 1768 TCTCCAGCAGGAAATGCCATTGCATGGACCAACAAATGCAATGCCATTTCCATTCC 1827  
DB 1176 TCTCCAGCAGGAAATGCCATTGCATGGACCAACAAATGCAATGCCATTTCCATTCC 1235  
QY 1828 TGTGCGCTCGAGACAGCCCGCTATGTGTCAACACCTATGAGAGCTACAGTGTCCGACCA 1887  
DB 1236 TGTGCGCTCGAGACAGCCCGCTATGTGTCAACACCTATGAGAGCTACAGTGTCCGACCA 1295  
QY 1888 ACAAGAAAGTGCAGTCCGGGCTACGAGCCCAACAGGATGGCACAGCTTGCCTGG 1941  
DB 1296 ACAAGAAAGTGCAGTCCGGGCTACGAGCCCAACAGGATGGCACAGCTTGCCTGG 1349

Search completed: August 9, 2004, 23:47:42  
Job time : 827 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 23:17:15 ; Search time 178 Seconds  
(without alignments)  
6787.236 Million cell updates/sec

Title: US-09-914-958B-35  
Perfect score: 2177  
Sequence: 1 cggaggctcgagccagcc.....tctcttgaaaaa 2177

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1314	60.4	1501	4	US-09-620-312D-1032
2	738	33.9	789	4	US-09-023-655-484
3	109	5.0	707	4	US-09-023-655-209
4	20	0.9	411	4	US-09-508-824-24
5	20	0.9	414	4	US-09-508-824-23
6	20	0.9	431	4	US-09-508-824-26
7	20	0.9	434	4	US-09-508-824-25
8	20	0.9	500	4	US-09-508-824-27
9	20	0.9	817	4	US-09-508-824-13
10	20	0.9	953	4	US-09-508-824-17
11	20	0.9	1001	4	US-09-508-824-18
12	20	0.9	1133	4	US-09-508-824-14
13	20	0.9	1134	4	US-09-508-824-20
14	20	0.9	1173	4	US-09-508-824-19
15	20	0.9	1185	4	US-09-508-824-15
16	20	0.9	1186	4	US-09-508-824-21
17	20	0.9	1916	4	US-09-508-824-16
18	20	0.9	1916	4	US-09-508-824-22
19	20	0.9	2332	4	US-09-508-824-2
20	20	0.9	4403765	3	US-09-103-840A-2
21	20	0.9	4411529	3	US-09-103-840A-1
22	19	0.9	299	4	US-09-132-316-11
23	19	0.9	299	4	US-09-132-316-45
24	19	0.9	326	4	US-09-132-316-44
25	19	0.9	403	3	US-09-328-111-360
26	19	0.9	422	4	US-09-132-316-25
27	19	0.9	461	4	US-09-132-316-20

28	19	0.9	642	4	US-09-252-991A-11468	Sequence 11468, A
29	19	0.9	747	4	US-09-252-991A-11489	Sequence 11489, A
30	19	0.9	768	4	US-09-235-451-6	Sequence 6, Appl
31	19	0.9	1281	3	US-09-105-537-19	Sequence 19, Appl
32	19	0.9	1512	4	US-09-199-637A-171	Sequence 171, Appl
33	19	0.9	1600	4	US-09-434-288-10	Sequence 10, Appl
34	19	0.9	1790	4	US-09-484-970B-67	Sequence 67, Appl
35	19	0.9	1977	4	US-09-548-372D-5	Sequence 5, Appl
36	19	0.9	1977	4	US-09-548-372D-5	Sequence 5, Appl
37	19	0.9	1977	4	US-09-551-853D-5	Sequence 5, Appl
38	19	0.9	2106	1	US-07-828-790B-1	Sequence 1, Appl
39	19	0.9	2184	4	US-09-484-970B-161	Sequence 161, Appl
40	19	0.9	2380	4	US-09-235-451-35	Sequence 35, Appl
41	19	0.9	2779	4	US-09-149-476-191	Sequence 191, Appl
42	19	0.9	2805	4	US-09-132-316-1	Sequence 1, Appl
43	19	0.9	2860	4	US-09-149-476-314	Sequence 314, Appl
44	19	0.9	2913	1	US-08-243-542-7	Sequence 7, Appl
45	19	0.9	2913	1	US-08-477-407-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1  
US-09-620-312D-1032  
; Sequence 1032, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_Fl\_genes Version 1.0  
; SEQ ID NO 1032  
; LENGTH: 1501  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (161)..(1468)  
US-09-620-312D-1032

Query Match 60.4%; Score 1314; DB 4; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1314; Conservative 0; Mismatches 0; Gaps 0;  
Qy 628 GTGGTGGCCAGCCTCTTTGCCGGCGCTCTGTGGCCTGTGGACAGAAAGGCTCTG 687  
Db 36 GTGGTGGCCAGCCTCTTTGCCGGCGCTCTGTGGCCTGTGGACAGAAAGGCTCTG 95





QY 1178 CTTTGACATGACAGAGCTGGAGATCTTTCTTCAACAAATTGGCTACCGAGCTCTC 1237  
| | | | |  
Db 121 CTTTGACATGACAGAGCTGGAGATCTTTCTTCAACAAATTGGCTACCGAGCTCTC 180  
| | | | |  
QY 1238 AGCCAAACCGCTCTTCGGCTCATCCGTAGAGACGAGACCCCTCATCGAGAGCT 1297  
| | | | |  
Db 181 AGCCAAACCGCTCTTCGGCTCATCCGTAGAGACGAGACCCCTCATCGAGAGCT 240  
| | | | |  
QY 1298 CAATCCCGGCGAGCTTTGGAGCTGAGGCCCGGGGACACAGGGGGTGTGTGACCGACTT 1357  
| | | | |  
Db 241 CAATCCCGGCGAGCTTTGGAGCTGAGGCCCGGGGACACAGGGGGTGTGTGACCGACTT 300  
| | | | |  
QY 1358 CGACGGAGACGGAGTGTGACCTCATCTTGTCCCATGGAGATCCATGCTCAGCCGCT 1417  
| | | | |  
Db 301 CGACGGAGACGGAGTGTGACCTCATCTTGTCCCATGGAGATCCATGCTCAGCCGCT 360  
| | | | |  
QY 1418 GTCCGCTTTCGGGGCAATCAGGCTTCAACAACTGGCTCGAGTGGTGCACCGCAC 1477  
| | | | |  
Db 361 GTCCGCTTTCGGGGCAATCAGGCTTCAACAACTGGCTCGAGTGGTGCACCGCAC 420  
| | | | |  
QY 1478 CCGGTTTGGGCTTTGGCAGGGAGCTAAGTCTGTCTACACCAAGAGAGTGGGC 1537  
| | | | |  
Db 421 CCGGTTTGGGCTTTGCCAGGGAGCTAAGTCTGTCTACACCAAGAGAGTGGGC 480  
| | | | |  
QY 1538 CCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATCGAGCCGTGGCACA 1597  
| | | | |  
Db 481 CCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATCGAGCCGTGGCACA 540  
| | | | |  
QY 1598 CTTTGGCTTGGGAAGGATGAAGCCAGCTGTGGAGTGCAGTGGCCAGATGGCAAGAT 1657  
| | | | |  
Db 541 CTTTGGCTTGGGAAGGATGAAGCCAGCTGTGGAGTGCAGTGGCCAGATGGCAAGAT 600  
| | | | |  
QY 1658 GGTGACCGGAACTGTGGCAGCGGGAGATGAATCAGTCTGGAGATCTTACCCCGG 1717  
| | | | |  
Db 601 GGTGACCGGAACTGTGGCAGCGGGAGATGAATCAGTCTGGAGATCTTACCCCGG 660  
| | | | |  
QY 1718 GGATGAGGACACACTTCAGACCCAGCCCACTGGAGTGGCCAGGATTTCCAGCA 1777  
| | | | |  
Db 661 GGATGAGGACACACTTCAGACCCAGCCCACTGGAGTGGCCAGGATTTCCAGCA 720  
| | | | |  
QY 1778 GGAATATGGCATTTGATGGACACCAATGAATGCATCCAGTTCCTTCCATTCGTGTGCCCTCG 1837  
| | | | |  
Db 721 GGAATATGGCATTTGATGGACACCAATGAATGCATCCAGTTCCTTCCATTCGTGTGCCCTCG 780  
| | | | |  
QY 1838 AGCAAGCC 1846  
| | | | |  
Db 781 AGCAAGCC 789  
| | | | |

RESULT 3  
US-09-023-655-209  
; Sequence 209, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 209:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 707 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: SYNORAB01  
; CLONE: 079378  
; US-09-023-655-209

Query Match 5.0%; Score 109; DB 4; Length 707;  
Best Local Similarity 100.0%; Pred. No. 4.8e-41;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1834 CTCGACAGACAGCCGCTATGTGTCAACACCTATCGAAGCTACAGTCCGACCAACA 1893  
| | | | |  
Db 1 CTCGACAGACAGCCGCTATGTGTCAACACCTATCGAAGCTACAGTCCGACCAACA 60  
| | | | |  
QY 1894 AGTGCAGTGGGCTACGAGCCCAACGAGATGGCACAGCTCGGTGG 1942  
| | | | |  
Db 61 AGTGCAGTGGGCTACGAGCCCAACGAGATGGCACAGCTCGGTGG 109  
| | | | |

RESULT 4  
US-09-508-824-24  
; Sequence 24, Application US/09508824  
; Patent No. 6635811  
; GENERAL INFORMATION:  
; APPLICANT: Flinham, John E  
; APPLICANT: Gale, Michael D  
; APPLICANT: Holdsworth, Michael J  
; TITLE OF INVENTION: Pre-harvest Sprouting  
; FILE REFERENCE: Mewburn  
; CURRENT APPLICATION NUMBER: US/09/508,824  
; CURRENT FILING DATE: 2000-05-17  
; PRIOR APPLICATION NUMBER: PCT/GB98/02835  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: GB 9720060.4  
; PRIOR FILING DATE: 1997-09-19  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 411  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; US-09-508-824-24

Query Match 0.9%; Score 20; DB 4; Length 411;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 45 GCGCGGCGCGCGCGCGCGCG 64  
| | | | |  
Db 15 GCGCGGCGCGCGCGCGCGCG 34  
| | | | |

RESULT 5  
US-09-508-824-23

```
; Sequence 23, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-23
```

```
Query Match 0.9%; Score 20; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 45 GCGCGCGCGCGCGCGCGCAG 64
|||||
Db 15 GCGCGCGCGCGCGCGCGCAG 34
```

## RESULT 6

```
US-09-508-824-26
; Sequence 26, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-26
```

```
Query Match 0.9%; Score 20; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 45 GCGCGCGCGCGCGCGCGCAG 64
|||||
Db 15 GCGCGCGCGCGCGCGCGCAG 34
```

## RESULT 7

```
US-09-508-824-25
; Sequence 25, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
```

```
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-25
```

```
Query Match 0.9%; Score 20; DB 4; Length 434;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 45 GCGCGCGCGCGCGCGCGCAG 64
|||||
Db 15 GCGCGCGCGCGCGCGCGCAG 34
```

## RESULT 8

```
US-09-508-824-27
; Sequence 27, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-27
```

```
Query Match 0.9%; Score 20; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 45 GCGCGCGCGCGCGCGCGCAG 64
|||||
Db 81 GCGCGCGCGCGCGCGCGCAG 100
```

## RESULT 9

```
US-09-508-824-13
; Sequence 13, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
```

```
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 817
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-13

Query Match          0.9%; Score 20; DB 4; Length 817;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GCGGCGCGCGCGCGCGCGAG 64
Db 106 GCGGCGCGCGCGCGCGCGAG 125

RESULT 10
US-09-508-824-17
; Sequence 17, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Newburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 953
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-17

Query Match          0.9%; Score 20; DB 4; Length 953;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GCGGCGCGCGCGCGCGCGAG 64
Db 201 GCGGCGCGCGCGCGCGCGAG 220

RESULT 11
US-09-508-824-18
; Sequence 18, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Newburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-18
```

```
Query Match          0.9%; Score 20; DB 4; Length 1001;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GCGGCGCGCGCGCGCGCGAG 64
Db 216 GCGGCGCGCGCGCGCGCGAG 235

RESULT 12
US-09-508-824-14
; Sequence 14, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Newburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-14

Query Match          0.9%; Score 20; DB 4; Length 1133;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GCGGCGCGCGCGCGCGCGAG 64
Db 369 GCGGCGCGCGCGCGCGCGAG 388

RESULT 13
US-09-508-824-20
; Sequence 20, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
; APPLICANT: Flintham, John E
; APPLICANT: Gale, Michael D
; APPLICANT: Holdsworth, Michael J
; TITLE OF INVENTION: Pre-harvest Sprouting
; FILE REFERENCE: Newburn
; CURRENT APPLICATION NUMBER: US/09/508,824
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/GB98/02835
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: GB 9720060.4
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-508-824-20

Query Match          0.9%; Score 20; DB 4; Length 1134;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GCGGCGCGCGCGCGCGCGAG 64
```

Db 369 GGCGGCGCGCGCGGCAG 388

RESULT 14

US-09-508-824-19  
 ; Sequence 19, Application US/09508824  
 ; Patent No. 6635811  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Flinham, John E  
 ; APPLICANT: Gale, Michael D  
 ; APPLICANT: Holdsworth, Michael J  
 ; TITLE OF INVENTION: Pre-harvest Sprouting  
 ; FILE REFERENCE: Mewburn  
 ; CURRENT APPLICATION NUMBER: US/09/508,824  
 ; CURRENT FILING DATE: 2000-05-17  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/02835  
 ; PRIOR FILING DATE: 1998-09-18  
 ; PRIOR APPLICATION NUMBER: GB 9720060.4  
 ; PRIOR FILING DATE: 1997-09-19  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 19  
 ; LENGTH: 1173  
 ; TYPE: DNA  
 ; ORGANISM: Triticum aestivum  
 US-09-508-824-19

Query Match 0.9%; Score 20; DB 4; Length 1173;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GGCGGCGCGCGCGGCAG 64

Db 428 GGCGGCGCGCGCGGCAG 447

RESULT 15

US-09-508-824-15  
 ; Sequence 15, Application US/09508824  
 ; Patent No. 6635811  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Flinham, John E  
 ; APPLICANT: Gale, Michael D  
 ; APPLICANT: Holdsworth, Michael J  
 ; TITLE OF INVENTION: Pre-harvest Sprouting  
 ; FILE REFERENCE: Mewburn  
 ; CURRENT APPLICATION NUMBER: US/09/508,824  
 ; CURRENT FILING DATE: 2000-05-17  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/02835  
 ; PRIOR FILING DATE: 1998-09-18  
 ; PRIOR APPLICATION NUMBER: GB 9720060.4  
 ; PRIOR FILING DATE: 1997-09-19  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 15  
 ; LENGTH: 1185  
 ; TYPE: DNA  
 ; ORGANISM: Triticum aestivum  
 US-09-508-824-15

Query Match 0.9%; Score 20; DB 4; Length 1185;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GGCGGCGCGCGCGGCAG 64

Db 419 GGCGGCGCGCGCGGCAG 438

Search completed: August 10, 2004, 03:35:34  
 Job time : 183 secs

1	1594	91.6	2039	12	US-10-332-929A-17	Sequence 17, Appl
2	1759	80.8	2589	12	US-10-332-929A-1	Sequence 1, Appl
3	1758	80.8	2039	12	US-10-332-929A-16	Sequence 16, Appl
4	1758	80.8	2145	12	US-10-332-929A-18	Sequence 18, Appl
5	1745	80.2	2507	16	US-10-295-027-182	Sequence 182, App
6	1745	80.2	2507	17	US-10-188-832-57	Sequence 57, Appl
7	1565	71.9	1849	9	US-09-822-830A-88	Sequence 88, Appl
8	1428	65.6	2263	9	US-09-765-231A-30	Sequence 30, Appl
9	1314	60.4	1501	15	US-10-037-270-1032	Sequence 1032, Ap
10	1314	60.4	1501	16	US-10-117-722-1032	Sequence 1032, Ap
11	1236	56.8	2178	16	US-10-295-027-180	Sequence 180, App
12	1236	56.8	2178	17	US-10-188-832-55	Sequence 55, Appl
13	738	33.9	789	17	US-10-641-643-484	Sequence 484, App
14	687	31.6	4794	16	US-10-295-027-184	Sequence 184, App

225 ACCATGATGGGCACCTTTGCGCATCGTCCGCCCGCCTACATCCACCGCACCGCCGCGCG

Query Match

Query Match

```
QY 373 AGTATGACCGGCGCCAGAAAGCGGCTGTGTGAATCATCGCGGTGATGAGCGGAGCTCACCCCT 432
Db 285 AGTATGACCGGCGCCAGAAAGCGGCTGTGTGAATCATCGCGGTGATGAGCGGAGCTCACCCCT 344
QY 433 ACTAGCGCTCGGGACCGGCGAGGGAGCGCATTTGGGGTCCACAGCTTCGACATCGACG 492
Db 345 ACTAGCGCTCGGGACCGGCGAGGGAGCGCATTTGGGGTCCACAGCTTCGACATCGACG 404
QY 493 GGGACGGCGGGAGGAGATCTACTCTTCAACACCAATTAATGCCCTTCTCGGGGGTGGCCA 552
Db 405 GGGACGGCGGGAGGAGATCTACTCTTCAACACCAATTAATGCCCTTCTCGGGGGTGGCCA 464
QY 553 CGTACACCGACAAGTTGTAAAGTTCCCGCAATTAACCGGTGGGAAGACATCTGAGCGATG 612
Db 465 CGTACACCGACAAGTTGTAAAGTTCCCGCAATTAACCGGTGGGAAGACATCTGAGCGATG 524
QY 613 AGGTCAACGTGGCCGCTGTGGCCAGCCTCTTTGCCGAGCGCTCTGTGGCCTGTGTGG 672
Db 525 AGGTCAACGTGGCCGCTGTGGCCAGCCTCTTTGCCGAGCGCTCTGTGGCCTGTGTGG 584
QY 673 ACAGAAAGGCTCTGACGCTACTCTATCTACATTTGCCAATTAACCGCTACGGTAATGTGG 732
Db 585 ACAGAAAGGCTCTGACGCTACTCTATCTACATTTGCCAATTAACCGCTACGGTAATGTGG 644
QY 733 GCCGTGATGCCCTCATTTGAATGAGCCCTGAGCCAGTGAACCTCTCCCGGGGCATTTCTGG 792
Db 645 GCCGTGATGCCCTCATTTGAATGAGCCCTGAGCCAGTGAACCTCTCCCGGGGCATTTCTGG 704
QY 793 CGCTCAGAGATGTGCTGCTGAGGCTGGGGTTCAGCAATATATACAGGGGGCCGAGCGTCA 852
Db 705 CGCTCAGAGATGTGCTGCTGAGGCTGGGGTTCAGCAATATATACAGGGGGCCGAGCGTCA 764
QY 853 GCGTGGGCCCCATCTCTACGACAGTGGCTTCGGATATCTTTCGCAATATGAGAAATGGGC 912
Db 765 GCGTGGGCCCCATCTCTACGACAGTGGCTTCGGATATCTTTCGCAATATGAGAAATGGGC 824
QY 913 CTAACCTCTTTTCCAAACCGGGGAGTGGACCTTTGTGACCGCTGGGGCAGTGTG 972
Db 825 CTAACCTCTTTTCCAAACCGGGGAGTGGACCTTTGTGACCGCTGGGGCAGTGTG 884
QY 973 GTGTGACGACCCCAACAGCATGGCGAGTGTGCGCCCTGCTGACTTCAACCGTGTG 1032
Db 885 GTGTGACGACCCCAACAGCATGGCGAGTGTGCGCCCTGCTGACTTCAACCGTGTG 944
QY 1033 GCAAAATGACATCTGTATGCAACTGGAATGGCCCCCAACCGCTCTATCTGCAAAATGA 1092
Db 945 GCAAAATGACATCTGTATGCAACTGGAATGGCCCCCAACCGCTCTATCTGCAAAATGA 1004
QY 1093 GCACCATGGGAAGTCCGCTTCCGGACATCGCCTCAACCCCAAGTTCTCCATGCCCTCCC 1152
Db 1005 GCACCATGGGAAGTCCGCTTCCGGACATCGCCTCAACCCCAAGTTCTCCATGCCCTCCC 1064
QY 1153 CTGTCCGACGCTCATCACCGCCGACTTTGACAAATGACCAGGAGCTGGAGATCTTCTTCA 1212
Db 1065 CTGTCCGACGCTCATCACCGCCGACTTTGACAAATGACCAGGAGCTGGAGATCTTCTTCA 1124
QY 1213 ACAACATTCCTAACCGACGCTCTCTAGCCAAACCGCCTTTCGCGGTCATCCGTAGAGAGC 1272
Db 1125 ACAACATTCCTAACCGACGCTCTCTAGCCAAACCGCCTTTCGCGGTCATCCGTAGAGAGC 1184
QY 1273 ACGGAGACCCCTCATTCGAGGAGCTCAATCCCGGCGACCGCTTGGAGCCTGAGGGCGGG 1332
Db 1185 ACGGAGACCCCTCATTCGAGGAGCTCAATCCCGGCGACCGCTTGGAGCCTGAGGGCGGG 1244
QY 1333 GCACAGGGGGTGTGTGACCGACTTCGACGAGACGGGATGTGGACCTCATCTTCTCCC 1392
Db 1245 GCACAGGGGGTGTGTGACCGACTTCGACGAGACGGGATGTGGACCTCATCTTCTCCC 1304
QY 1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCATCAGGGCTTCAACAA 1452
Db 1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCATCAGGGCTTCAACAA 1364
```

```
QY 1453 ACTGCTCGAGTGTGTGCCACGACCCCGTTTGGGCGCTTTGCCAGGGAGCTAAGGTGG 1512
Db 1365 ACTGCTCGAGTGTGTGCCACGACCCCGTTTGGGCGCTTTGCCAGGGAGCTAAGGTGG 1424
QY 1513 TGCTCTACACCAAGAGAGTGGGCGCCACCTTGAGGATCATCGACGGGGCTCAGGCTACC 1572
Db 1425 TGCTCTACACCAAGAGAGTGGGCGCCACCTTGAGGATCATCGACGGGGCTCAGGCTACC 1484
QY 1573 TGTGTGAGATGGAGCCCGTGGCAGACATTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
Db 1485 TGTGTGAGATGGAGCCCGTGGCAGACATTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544
QY 1633 AGGTGACCTGGCAGATGGCAAGATGTGTAGCCGGAACGTGGCCACGCGGGAGATGAAC 1692
Db 1545 AGGTGACCTGGCAGATGGCAAGATGTGTAGCCGGAACGTGGCCACGCGGGAGATGAAC 1604
QY 1693 CAGTGTGAGATGCTCTTACCCCGGGGATGAGGACACACTTCAGGACCCAGCCCACTGG 1752
Db 1605 CAGTGTGAGATGCTCTTACCCCGGGGATGAGGACACACTTCAGGACCCAGCCCACTGG 1664
QY 1753 AGTGTGGCCAAAGGATTTCTCCAGCAGAGAAATGGCCATTGTCATGACACCAATGAATGCA 1812
Db 1665 AGTGTGGCCAAAGGATTTCTCCAGCAGAGAAATGGCCATTGTCATGACACCAATGAATGCA 1724
QY 1813 TCCAGTTCCTCCATTTCTGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACTATGGAAGCT 1872
Db 1725 TCCAGTTCCTCCATTTCTGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACTATGGAAGCT 1784
QY 1873 ACAGTGTGGGACCAACAAAGAGTGCAGTGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
Db 1785 ACAGTGTGGGACCAACAAAGAGTGCAGTGGGGCTACGAGCCCAACGAGGATGGCACAG 1844
QY 1933 CCGTGTGGGCTGTGGAGCCCTGTGTGAAGATAGTGACACCAACAAAGTTGGGAAGAGCC 1992
Db 1845 CCGTGTGGGCTGTGGAGCCCTGTGTGAAGATAGTGACACCAACAAAGTTGGGAAGAGCC 1904
QY 1993 TTGGTCCCTGATCACTCAATCACTGCTTGAATCACCGCTTGAATCACCGCTTGAATCAC 2052
Db 1905 TTGGTCCCTGATCACTCAATCACTGCTTGAATCACCGCTTGAATCACCGCTTGAATCAC 1964
QY 2053 GAACACTTACCTGGAACTTCACTGAGCAGGATACAACTTCTATTGTATTAGCTATTAA 2112
Db 1965 GAACACTTACCTGGAACTTCACTGAGCAGGATACAACTTCTATTGTATTAGCTATTAA 2024
QY 2113 TACATTAAGATTTGGGGTGTACCTTACATAATAATTTCCCATTTCCCTTTGAAAAAAA 2172
Db 2025 TACATTAAGATTTGGGGTGTACCTTACATAATAATTTCCCATTTCCCTTTGAAAAAAA 2084
QY 2173 AAAAA 2177
Db 2085 AAAAA 2089
```

## RESULT 2

```
US-10-332-929A-1
; Sequence 1, Application US/10332929A
; Publication No. US2004007286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2589
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2498)..(2498)
; OTHER INFORMATION: n is a, c, g, or t
```







[illegible]

```

RESULT 5
US-70-295-027-182
; Sequence 182, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Haynes, Richard
; APPLICANT: Havezzi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard

```













1339 GGGGTGGTGACCGACTTCGACGGAGCGGATGCTGGACCTCATCTGTCCCATGGAG 1398  
1473 GGGGTGGTGACCGACTTCGACGGAGCGGATGCTGGACCTCATCTGTCCCATGGAG 1532  
1399 AGTCATGGCTCAGCGCTGTCGGTCTTCGGGGCAATCAGGGCTTCAACAACAACCTGGC 1458  
1533 AGTCATGGCTCAGCGCTGTCGGTCTTCGGGGCAATCAGGGCTTCAACAACAACCTGGC 1592  
1459 TGCAGTGGTGCC-ACGACCCCGGTTTGGGACCTTTCGACGGGAGCTTAAGTGTGCTC 1517  
1593 TGCAGTGGTGCCAAACGACCCGGTTTGGGACCTTTCGACGGGAGCTTAAGTGTGCTC 1652  
1518 TACACCAAGAAGTGGGGCCCACTCAGGATCATCGAGGGGCTCAGGCTACCTGTGT 1577  
1653 TACACCAAGAAGTGGGGCCCACTCAGGATCATCGAGGGGCTCAGGCTACCTGTGT 1712  
1578 GAGATGGAGCCGCTGGCACACTTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637  
1713 GAGATGGAGCCGCTGGCACACTTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772  
1638 ACCTGGCCAGATGGCAAGATGGTGAAGCCGGAACGTTGGCCAGCGGGGAGTGAATCACTG 1697  
1773 ACCTGGCCAGATGGCAAGATGGTGAAGCCGGAACGTTGGCCAGCGGGGAGTGAATCACTG 1832  
1698 CTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGGTG 1757  
1833 CTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGGTG 1892  
1758 GSCCAAGGATCTCCAGCAGGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816  
1893 GSCCAAGGATCTCCAGCAGGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1952  
1817 GTTCCCATTCGTTGCTCGAGACAAGCCCGTATGTGTCAACACTATGGAAGCTACAG 1876  
1953 GTTCCCATTCGTTGCTCGAGACAAGCCCGTATGTGTCAACACTATGGAAGCTACAG 2012  
1877 GTGCCGGACCAAGAGTGCAGTCGGGCTACAGGCCCAACAGGAGTGGCAGACGCTG 1936  
2013 GTGCCGGACCAAGAGTGCAGTCGGGCTACAGGCCCAACAGGAGTGGCAGACGCTG 2072  
1937 CGTGG 1941  
2073 CGTGG 2077

RESULT 9  
US-10-037-270-1032  
Sequence 1032, Application US/10037270  
Publication No. US20030104529A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yungting  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillingshast, John  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04

PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt FL genes Version 1.0  
SEQ ID NO 1032  
LENGTH: 1501  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (161)..(1468)  
US-10-037-270-1032  
Query Match 60.4%; Score 1314; DB 15; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 628 GTGGTGTGGCCAGCCCTCTTTGCCGAGCGCTCTGTGGCCTGTGTGGACAGAAAGGCTCTG 687  
Db 36 GTGGTGTGGCCAGCCCTCTTTGCCGAGCGCTCTGTGGCCTGTGTGGACAGAAAGGCTCTG 95  
QY 688 GACGCTACTCTATCTACATTTGCCAATTTAGCCTACGTAATTTGGCCCTGATGCCCTCA 747  
Db 96 GACGCTACTCTATCTACATTTGCCAATTTAGCCTACGTAATTTGGCCCTGATGCCCTCA 155  
QY 748 TTGAATAGACCCCTCAGGCCAGTACCTCTCCCGGGGCAATCTGTGGCTCAGAGATGTGG 807  
Db 156 TTGAATAGACCCCTCAGGCCAGTACCTCTCCCGGGGCAATCTGTGGCTCAGAGATGTGG 215  
QY 808 CTGCTGAGGCTGGGCTCAGCAAAATATACAGGGGGCGAGGCGTCAGCGTGGGCCCATCC 867  
Db 216 CTGCTGAGGCTGGGCTCAGCAAAATATACAGGGGGCGAGGCGTCAGCGTGGGCCCATCC 275  
QY 868 TCAGCAGAGTGCCTCGGATATCTTCTGGGACAAATGAGATGGGCTACTTCTCTTTCC 927  
Db 276 TCAGCAGAGTGCCTCGGATATCTTCTGGGACAAATGAGATGGGCTACTTCTCTTTCC 335  
QY 928 ACAACCGGGCGCATGCGACCTTTTGGAGCGCTGGGCCAGTGTGTGTGGAGCAGCCCC 987  
Db 336 ACAACCGGGCGCATGCGACCTTTTGGAGCGCTGGGCCAGTGTGTGTGGAGCAGCCCC 395  
QY 988 ACCAGCATGGCGAGTGTGCGCTCTGGCTGACTTCAACCGTGAATGGCAAGTGAATCG 1047  
Db 396 ACCAGCATGGCGAGTGTGCGCTCTGGCTGACTTCAACCGTGAATGGCAAGTGAATCG 455  
QY 1048 TCTATGGCAACTGGAATGGCCCCCAGCGCTCTATCTGCAATGAGCACCCATGGGAAG 1107  
Db 456 TCTATGGCAACTGGAATGGCCCCCAGCGCTCTATCTGCAATGAGCACCCATGGGAAG 515  
QY 1108 TCGGCTTCGGGACATCGGCTCACCAGTGTCTCCATGCCCTCCCTGTCCGACCGTCA 1167  
Db 516 TCGGCTTCGGGACATCGGCTCACCAGTGTCTCCATGCCCTCCCTGTCCGACCGTCA 575  
QY 1168 TCACCGCCGACTTTGACAAATGACAGGAGCTGGAGATCTTCTTCAACAACATTCCTACC 1227  
Db 576 TCACCGCCGACTTTGACAAATGACAGGAGCTGGAGATCTTCTTCAACAACATTCCTACC 635  
QY 1228 GCAGCTCTCAGCAACCGCTCTTCCGCGTATCCGATAGAGAGCAGGAGCCCCCTCA 1287  
Db 636 GCAGCTCTCAGCAACCGCTCTTCCGCGTATCCGATAGAGAGCAGGAGCCCCCTCA 695  
QY 1288 TCGAGGAGCTCAATCCCGCGCAGCGCTTGGAGCCTGAGGGCCGAGGAGTGTGG 1347  
Db 696 TCGAGGAGCTCAATCCCGCGCAGCGCTTGGAGCCTGAGGGCCGAGGAGTGTGG 755  
QY 1348 TGACCGACTTCGACGAGAGCGGAGTGTGACCTCATCTTGTCCATGAGAGTCCATGG 1407  
Db 756 TGACCGACTTCGACGAGAGCGGAGTGTGACCTCATCTTGTCCATGAGAGTCCATGG 815  
QY 1408 CTCAGCGCTGTCCGTCCTCCGGGCAATCAGGGCTTCAACAACACTGCTGCGAGTGG 1467

Db 816 CTGAGCGCTGTCGGTCTTCGGGGCAATCAGGGCTTCAACAACAACACTGCGTGGAGTGG 875  
QY 1468 TGCACAGCACCCCGTTTGGGCGCTTTCACAGGGAGCTAAGGTCGTCTCTACACCAAGA 1527  
Db 876 TGCACAGCACCCCGTTTGGGCGCTTTCACAGGGAGCTAAGGTCGTCTCTACACCAAGA 935  
QY 1528 AGAGTGGGCGCCACCTGAGGATCATCAGCGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587  
Db 936 AGAGTGGGCGCCACCTGAGGATCATCAGCGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995  
QY 1588 CCGTGGCACACTTTGGGCTCGGGAAGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1647  
Db 996 CCGTGGCACACTTTGGGCTCGGGAAGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1055  
QY 1648 ATGGCAAGATGGTGAAGCGGAAAGTGGCCAGCGGGGAGATGAATCAGTGTGTGAGATGCC 1707  
Db 1056 ATGGCAAGATGGTGAAGCGGAAAGTGGCCAGCGGGGAGATGAATCAGTGTGTGAGATGCC 1115  
QY 1708 TCTACCCCGGGATGAGGACACACTTCAGACCCAGCCCTGAGGTGTGGCCAGGAT 1767  
Db 1116 TCTACCCCGGGATGAGGACACACTTCAGACCCAGCCCTGAGGTGTGGCCAGGAT 1175  
QY 1768 TCTCCACGAGGAAATGGCCATTTGCATGACACCAATGAATGCATCCAGTTCCTCATTCG 1827  
Db 1176 TCTCCACGAGGAAATGGCCATTTGCATGACACCAATGAATGCATCCAGTTCCTCATTCG 1235  
QY 1828 TGTGCCCTCGAGACAGCCCGTATGTCTCAACCTTATGGAAGCTACAGGTGCGGACCA 1887  
Db 1236 TGTGCCCTCGAGACAGCCCGTATGTCTCAACCTTATGGAAGCTACAGGTGCGGACCA 1295  
QY 1888 ACAAGAAGTGCAGTCGGGGCTACAGCCCAACAGAGATGCGACAGCTGCGTGG 1941  
Db 1296 ACAAGAAGTGCAGTCGGGGCTACAGCCCAACAGAGATGCGACAGCTGCGTGG 1349

RESULT 10

US-10-117-722-1032  
; Sequence 1032, Application US/10117722  
; Publication No. US20030219744A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
; FILE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2BCIP  
; CURRENT APPLICATION NUMBER: US/10/117,722  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 1032  
; LENGTH: 1501  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (161)..(1468)  
US-10-117-722-1032  
Query Match 60.4%; Score 1314; DB 16; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 628 GTGGTGGCCGCTTTGGCGAGCGCTCTGTGCGCTGTGTGGACAGAAAGGCTCTG 687

Db 36 GTGGTGGCCAGCCCTTTTGGCGGACGCTCTGTGGCGCTGTGTGGACAGAAAGGCTCTG 95  
QY 688 GAGCTACTCTATCTACTATTCGCAATTTACGCCCTACGTTAATGTGGCGCTGTGATCCCTCA 747  
Db 96 GAGCTACTCTATCTACTATTCGCAATTTACGCCCTACGTTAATGTGGCGCTGTGATCCCTCA 155  
QY 748 TTGAATGGACCTGAGGCGAGTGAACCTTCCGGGGCATTTCTGGCGCTCAGAGATGG 807  
Db 156 TTGAATGGACCTGAGGCGAGTGAACCTTCCGGGGCATTTCTGGCGCTCAGAGATGG 215  
QY 808 CTGCTGAGGCTGGGCTCAGCAATATATACAGGGGCGAGGCGTACGCTGGGCGCCCATCC 867  
Db 216 CTGCTGAGGCTGGGCTCAGCAATATATACAGGGGCGAGGCGTACGCTGGGCGCCCATCC 275  
QY 868 TCAGCAGCAGTGTCTCGGATATCTTGTGCAATATGAGATGGGCTTAACCTTCTCTTTTCC 927  
Db 276 TCAGCAGCAGTGTCTCGGATATCTTGTGCAATATGAGATGGGCTTAACCTTCTCTTTTCC 335  
QY 928 ACAACGGGCGATGGCACCTTTGTGACGCTCGGCGAGTGTGTGGACGACGACCC 987  
Db 336 ACAACGGGCGATGGCACCTTTGTGACGCTCGGCGAGTGTGTGGACGACGACCC 395  
QY 988 ACCAGCATGGGCGAGTGTGCGGCTGCTGACTTCAACGCTGATGGCAAGTGGACATCG 1047  
Db 396 ACCAGCATGGGCGAGTGTGCGGCTGCTGACTTCAACGCTGATGGCAAGTGGACATCG 455  
QY 1048 TCTATGGCAACTTGAATGGCCCCCACCGCTCTATCTGCAATATGAGCACCCATGGGAAG 1107  
Db 456 TCTATGGCAACTTGAATGGCCCCCACCGCTCTATCTGCAATATGAGCACCCATGGGAAG 515  
QY 1108 TCGCTTCCGGGACATGCGCTCACCAAGTTCTCCATGCGCCTCCCTGTCCGACGCTCA 1167  
Db 516 TCGCTTCCGGGACATGCGCTCACCAAGTTCTCCATGCGCCTCCCTGTCCGACGCTCA 575  
QY 1168 TCACCGCGCACTTTGACATGACAGGCTGGAGATCTTCTTCAACAACATTTGCTTACC 1227  
Db 576 TCACCGCGCACTTTGACATGACAGGCTGGAGATCTTCTTCAACAACATTTGCTTACC 635  
QY 1228 GCAGCTCTCAGCAACCGCTCTTCCGCGTCAATCGTAGAGAGCAGGAGACCCCTCA 1287  
Db 636 GCAGCTCTCAGCAACCGCTCTTCCGCGTCAATCGTAGAGAGCAGGAGACCCCTCA 695  
QY 1288 TCAGGAGCTCAATCCCGGCGACGCTTGGAGCTCAGGCGCGGGGACAGGGGGTGTGG 1347  
Db 696 TCAGGAGCTCAATCCCGGCGACGCTTGGAGCTCAGGCGCGGGGACAGGGGGTGTGG 755  
QY 1348 TGACCGACTTCGACGAGGAGGATGCTGACCTCATCTTGTCCCATGGAGAGTCCATGG 1407  
Db 756 TGACCGACTTCGACGAGGAGGATGCTGACCTCATCTTGTCCCATGGAGAGTCCATGG 815  
QY 1408 CTCAGCGCTGTCCGCTTTCGGGGCAATCAGGGCTTCAACAACAACCTGGCTCGAGTGG 1467  
Db 816 CTCAGCGCTGTCCGCTTTCGGGGCAATCAGGGCTTCAACAACAACCTGGCTCGAGTGG 875  
QY 1468 TGCACGACCCCGTTTGGGCGCTTTGCCAGGGGAGCTAAGGTCGTCTCAACCAAGA 1527  
Db 876 TGCACGACCCCGTTTGGGCGCTTTGCCAGGGGAGCTAAGGTCGTCTCAACCAAGA 935  
QY 1528 AGAGTGGGCGCCACCTGAGGATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587  
Db 936 AGAGTGGGCGCCACCTGAGGATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995  
QY 1588 CCGTGGCACACTTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1647  
Db 996 CCGTGGCACACTTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1055  
QY 1648 ATGGCAAGATGGTGAAGCGGAAAGTGGCCAGCGGGGAGATGAATCAGTGTGTGAGATGCC 1707  
Db 1056 ATGGCAAGATGGTGAAGCGGAAAGTGGCCAGCGGGGAGATGAATCAGTGTGTGAGATGCC 1115  
QY 1708 TCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCTGAGGTGTGGCCAGGAT 1767  
Db 1116 TCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCTGAGGTGTGGCCAGGAT 1175

```
QY 1768 TCTCCACAGCAAAATGCCATTGTCATGGACACCAATCAATGATCCAGTTCCTCCATTGC 1827
Db 1176 TCTCCACAGCAAAATGCCATTGTCATGGACACCAATCAATGATCCAGTTCCTCCATTGC 1235
QY 1828 TGTGCGCTCGAGACAAGCCGCTATGTGTCAACACCTATGGAAGCTACAGAGTGCAGGACCA 1887
Db 1236 TGTGCGCTCGAGACAAGCCGCTATGTGTCAACACCTATGGAAGCTACAGAGTGCAGGACCA 1295
QY 1888 ACAAGAAGTCAGTGGGGCTACGAGGCCCAACAGAGATGGCACACCTGCGGTGG 1941
Db 1296 ACAAGAAGTCAGTGGGGCTACGAGGCCCAACAGAGATGGCACACCTGCGGTGG 1349

RESULT 11
US-10-295-027-180
; Sequence 180, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 180
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-180

Query Match 56.8%; Score 1236; DB 16; Length 2178;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 468 GGGGTCAACGCTCGACATCGACGGGACGCGGGAGGAGATCTACTTCTCTCAACACC 527
Db 34 GGGGTCAACGCTCGACATCGACGGGACGCGGGAGGAGATCTACTTCTCTCAACACC 93
QY 528 AATAATGCCTTCTCGGGGGTGGCCAGCTACCGACAAGTTGTTCAAGTTCCGCAATAAC 587
```

```
QY 1668 AACGTGCGCAGCGGGGAGATGAACCTAGTGTGAGATCCTCTACCCCGGGATGAGGAC 1727
Db 1234 AACGTGCGCAGCGGGGAGATGAACCTAGTGTGAGATCCTCTACCCCGGGATGAGGAC 1293
QY 1728 ACACCTTCAGGACCCAGCCCACTGGAG 1754
Db 1294 ACACCTTCAGGACCCAGCCCACTGGAG 1320

RESULT 12
US-10-188-832-55
; Sequence 55, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: And Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002330US
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-188-832-55

Query Match 56.8%; Score 1236; DB 17; Length 2178;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 468 GGGGTACAGCCTCGACATCGACGGGGAGCGCGGGGAGAGATCTACTTCTCAACACC 527
Db 34 GGGGTACAGCCTCGACATCGACGGGGAGCGCGGGGAGAGATCTACTTCTCAACACC 93
QY 528 AATAATGCCCTTCGGGGGTGGCCAGTACCGCAAGTGTTCAGTTTCGCAATAAC 587
Db 94 AATAATGCCCTTCGGGGGTGGCCAGTACCGCAAGTGTTCAGTTTCGCAATAAC 153
QY 588 CGGTGGGAAGACATCTGAGCGATGAGTCAAGTGGCCCGCTGGTGGCCAGCCTCTTT 647
Db 154 CGGTGGGAAGACATCTGAGCGATGAGTCAAGTGGCCCGCTGGTGGCCAGCCTCTTT 213
QY 648 GCGGACGCTCTGTGGCCCTGTGGACAGAAAGGCTCTGGAGCGTACTTATCTACATT 707
Db 214 GCGGACGCTCTGTGGCCCTGTGGACAGAAAGGCTCTGGAGCGTACTTATCTACATT 273
QY 708 GCGAATTACGCTACGGTAATGTGGCCCTGTGATGCCCTCATTTGAATGACCCCTGAGGCC 767
Db 274 GCGAATTACGCTACGGTAATGTGGCCCTGTGATGCCCTCATTTGAATGACCCCTGAGGCC 333
QY 768 AGTGACCTCTCCCGGGGATCTTGGCGCTCAGAGATGTGGCTGCTGAGCTGGGTGAGC 827
Db 334 AGTGACCTCTCCCGGGGATCTTGGCGCTCAGAGATGTGGCTGCTGAGCTGGGTGAGC 393
QY 828 AAATATACAGGGGGCGAGGGCTCAGCGTGGGCCCCATCTCTCAGCAGCAGTGCCTCGGAT 887
Db 394 AAATATACAGGGGGCGAGGGCTCAGCGTGGGCCCCATCTCTCAGCAGCAGTGCCTCGGAT 453
```

```
QY 888 ATCTTCTGCACAAATGAGAAATGGCCCTAACTTCTCTTTTCCAAACCCGGGGATGGCACC 947
Db 454 ATCTTCTGCACAAATGAGAAATGGCCCTAACTTCTCTTTTCCAAACCCGGGGATGGCACC 513
QY 948 TTTGTGAGACCTGCGGCCAGTGTGTGTGGAGAGACCCCAACAGCATGGCGAGGTGTC 1007
Db 514 TTTGTGAGACCTGCGGCCAGTGTGTGTGGAGAGACCCCAACAGCATGGCGAGGTGTC 573
QY 1008 GCCTCGCTGACTTCAACCGTGTATGCATAATGGAGATCGTCTATGCAATGAAATGGC 1067
Db 574 GCCTCGCTGACTTCAACCGTGTATGCATAATGGAGATCGTCTATGCAATGAAATGGC 633
QY 1068 CCCACCGCCTCTATCTGCAATGAGCACCCATGGGAAGTCCGCTTCCGGGACATCGCC 1127
Db 634 CCCACCGCCTCTATCTGCAATGAGCACCCATGGGAAGTCCGCTTCCGGGACATCGCC 693
QY 1128 TCACCAAGTTCATGCGCTCCCTGTCCGCAACCGGTATACACCGCGACTTTGACAAT 1187
Db 694 TCACCAAGTTCATGCGCTCCCTGTCCGCAACCGGTATACACCGCGACTTTGACAAT 753
QY 1188 GACGAGAGCTGAGATCTTCTTCAACAAATTCGCTACCGAGCTCTCAGCCCAACCGC 1247
Db 754 GACGAGAGCTGAGATCTTCTTCAACAAATTCGCTACCGAGCTCTCAGCCCAACCGC 813
QY 1248 CTCTTCGCGTCACTCGTATAGAGACGAGACCCCTCTCATCGAGGAGCTCAATCCCGGC 1307
Db 814 CTCTTCGCGTCACTCGTATAGAGACGAGACCCCTCTCATCGAGGAGCTCAATCCCGGC 873
QY 1308 GAGCCCTTGAGCCTGAGGGCCGGGGACACAGGGGTGTGTGACCGATTTCGACGAGAC 1367
Db 874 GAGCCCTTGAGCCTGAGGGCCGGGGACACAGGGGTGTGTGACCGATTTCGACGAGAC 933
QY 1368 GGGATGCTGAGCCTCATCTTGTCCATGAGAGTCCATGSCCTCAGCGCTGTCCGCTTC 1427
Db 934 GGGATGCTGAGCCTCATCTTGTCCATGAGAGTCCATGSCCTCAGCGCTGTCCGCTTC 993
QY 1428 CGGGGCAATCAGGGCTTCAACAACTGGCTCGAGTGTGCCAGCACCCCGTTTGGG 1487
Db 994 CGGGGCAATCAGGGCTTCAACAACTGGCTCGAGTGTGCCAGCACCCCGTTTGGG 1053
QY 1488 GCCTTTGCCAGGGGAGCTAAGTGTGCTCTACCAAGAGAGTGGGCCCCACCTGAGG 1547
Db 1054 GCCTTTGCCAGGGGAGCTAAGTGTGCTCTACCAAGAGAGTGGGCCCCACCTGAGG 1113
QY 1548 ATCATCGACGGGGCTCAGGCTACCTGTGTGAGTGGAGCCGTGGCACACTTTGGCCTG 1607
Db 1114 ATCATCGACGGGGCTCAGGCTACCTGTGTGAGTGGAGCCCGTGGCACACTTTGGCCTG 1173
QY 1608 GGGAAAGGATGAAGCCAGCAGTGTGGAGTGAAGTGGCCAGATGGCAAGATGGTGAAGCCG 1667
Db 1174 GGGAAAGGATGAAGCCAGCAGTGTGGAGTGAAGTGGCCAGATGGCAAGATGGTGAAGCCG 1233
QY 1668 AACGTGCGCAGCGGGGAGATGAACCTAGTGTGAGATCCTCTACCCCGGGATGAGGAC 1727
Db 1234 AACGTGCGCAGCGGGGAGATGAACCTAGTGTGAGATCCTCTACCCCGGGATGAGGAC 1293
QY 1728 ACACCTTCAGGACCCAGCCCACTGGAG 1754
Db 1294 ACACCTTCAGGACCCAGCCCACTGGAG 1320
```

```
RESULT 13
US-10-641-643-484
; Sequence 484, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; NUMBER OF SEQUENCES: 1508
```

;;  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
;; STREET: 3174 PORTER DRIVE  
;; CITY: PALO ALTO  
;; STATE: CALIFORNIA  
;; COUNTRY: USA  
;; ZIP: 94304  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/641,643  
;; FILING DATE: 14-Aug-2003  
;; CLASSIFICATION: <Unknown>  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: <Unknown>  
;; FILING DATE: <Unknown>  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Zeller, Karen J.  
;; REGISTRATION NUMBER: 37,071  
;; REFERENCE/DOCKET NUMBER: PA-0001 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (650) 855-0555  
;; TELEFAX: (650) 845-4166  
;;  
;; INFORMATION FOR SEQ ID NO: 484:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 789 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: SYNORAB01  
;; CLONE: 192279  
;;  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 484 :  
US-10-641-643-484

Query Match 33.9%; Score 738; DB 17; Length 789;

Best Local Similarity 99.9%; Pred.No. 0; Mismatches 1; Indels 0; Gaps 0;  
Matches 788; Conservative 0;

QY	1058	CTGGAAATGCCCCCAGCGCTCTATCTGCAATGAGCAGCCCATGGGAAGTCCGTTCCG	1117
DB	1	CTGGAAATGCCCCCAGCGCTCTATCTGCAATGAGCAGCCCATGGGAAGTCCGTTCCG	60
QY	1118	GGACATGCGCTCACCACAGTTCCATGCTCCCTGTCGCGACGGTCAATCAGCGCGA	1177
DB	61	GGACATGCGCTCACCACAGTTCCATGCTCCCTGTCGCGACGGTCAATCAGCGCGA	120
QY	1178	CTTTGACATGACGAGGCTGGAGATCTTCTTCAACACATTGCTACCGAGCTCCTC	1237
DB	121	CTTTGACATGACGAGGCTGGAGATCTTCTTCAACACATTGCTACCGAGCTCCTC	180
QY	1238	AGCCAAACCGCTCTTCCGCGTTCATCGTAGAGCAGGAGACCCCTCATCGAGAGCT	1297
DB	181	AGCCAAACCGCTCTTCCGCGTTCATCGTAGAGCAGGAGACCCCTCATCGAGAGCT	240
QY	1298	CAATCCCGCGACGCTTGGAGCTTGAGCGCGGCGCAAGGGGTTGGTGAACGACTT	1357
DB	241	CAATCCCGCGACGCTTGGAGCTTGAGCGCGGCGCAAGGGGTTGGTGAACGACTT	300
QY	1358	CGACGAGAGCGGATGCTGGACTCATCTTGTCCCATGAGAGTCCATGCTCAGCGCT	1417
DB	301	CGACGAGAGCGGATGCTGGACTCATCTTGTCCCATGAGAGTCCATGCTCAGCGCT	360
QY	1418	GTCCGCTTTCCGGGGCAATCAGGGCTTCAACAACTCGGCTCGAGTGGTCCAGCGAC	1477
DB	361	GTCCGCTTTCCGGGGCAATCAGGGCTTCAACAACTCGGCTCGAGTGGTCCAGCGAC	420
QY	1478	CGGTTTGGGGCTTTTCCAGGGGAGCTAAGTCTGTGCTCTACACCAAGAGTGGGG	1537
DB	421	CGGTTTGGGGCTTTTCCAGGGGAGCTAAGTCTGTGCTCTACACCAAGAGTGGGG	480

QY	1538	CCACTGAGGATCATCGACGGGGCTCAGGCTACTGTGTGAGATGGAGCCCGTGGCACA	1597
DB	481	CCACTGAGGATCATCGACGGGGCTCAGGCTACTGTGTGAGATGGAGCCCGTGGCACA	540
QY	1598	CTTTGGCTGGGGAAGGATGAAGCAGCAGTGTGGAGTGACGTGGCCAGATGGCAAGAT	1657
DB	541	CTTTGGCTGGGGAAGGATGAAGCAGCAGTGTGGAGTGACGTGGCCAGATGGCAAGAT	600
QY	1658	GCTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATCTCTACCCCG	1717
DB	601	GCTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATCTCTACCCCG	660
QY	1718	GGATGAGGACACACTTTCAGGACCCAGCCCCCACTGGAGTGTGGCCAAAGGATTCCTCCAGCA	1777
DB	661	GGATGAGGACACACTTTCAGGACCCAGCCCCCACTGGAGTGTGGCCAAAGGATTCCTCCAGCA	720
QY	1778	GGAAATGCCCATTCGATGCGACACCAATGAATGAATCCAGTTCCTCCTGCTGCTCG	1837
DB	721	GGAAATGCCCATTCGATGCGACACCAATGAATGAATCCAGTTCCTCCTGCTGCTCG	780
QY	1838	AGACAAGCC 1846	
DB	781	AGACAAGCC 789	

## RESULT 14

US-10-295-027-184  
; Sequence 184, Application US/10295027  
; Publication No. US20030232350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and  
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer  
; FILE REFERENCE: 018501-C12500US  
; CURRENT APPLICATION NUMBER: US/10/295,027  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10  
; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; PRIOR FILING DATE: 2002-02-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 184  
; LENGTH: 4794  
; TYPE: DNA  
; ORGANISM: Homo sapiens



**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 9, 2004, 23:06:56 ; Search time 5087 Seconds  
(without alignments)  
12779.628 Million cell updates/sec

Title: US-09-914-958B-35  
Perfect score: 2177  
Sequence: 1 cggagagctcgagccagcc.....tcctcttgaaaaaaaaaa 2177

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1438	66.1	2147	11	BC042687
2	719	33.0	781	12	B1550776
3	653	30.0	841	12	BG708845
4	632	29.0	656	13	BQ183029

5	626	28.8	636	10	BF727258
6	613	28.2	1101	13	BQ067249
7	590	27.1	770	12	BF758703
8	566	26.0	926	10	BF527664
9	553	25.4	904	13	BA451456
10	519	23.8	735	12	BG911024
11	516	23.7	567	12	BM707643
12	514	23.6	1063	12	B1596960
13	507	23.3	578	14	CD676152
14	501	23.0	804	12	B1458314
15	496	22.8	608	9	AU123536
16	488	22.4	540	14	CB153273
17	486	22.3	598	13	BQ637467
18	483	22.2	566	12	BG721807
19	477	21.9	511	9	AI791928
20	473	21.7	551	12	BM699706
21	466	21.4	489	10	BF725127
22	458	21.0	580	12	B1549399
23	455	20.9	1022	13	BA460454
24	451	20.7	713	9	AI792073
25	447	20.5	539	14	CB152752
26	442	20.3	477	13	BX102537
27	411	18.9	622	12	B1824952
28	409	18.8	622	12	BG716170
29	399	18.3	399	14	CA388867
30	394	18.1	549	12	BM706915
31	383	17.6	533	10	BE503986
32	369	16.9	430	10	BF930592
33	357	16.4	508	10	BF088585
34	348	16.0	433	12	BG900984
35	337	15.5	860	12	B1819378
36	327	15.0	438	12	BG999412
37	326	15.0	338	10	BF727076
38	313	14.4	398	12	BG989358
39	305	14.0	470	10	BF727075
40	305	14.0	729	12	BG818593
41	295	13.6	350	10	BF736863
42	272	12.5	470	10	BE503764
43	237	10.9	539	10	BF849175
44	226	10.4	418	14	T77333
45	218	10.0	380	12	BG920409

ALIGNMENTS

RESULT 1

BC042687

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

BC042687 Mus musculus, clone IMAGE:3157049, mRNA. 2147 bp linear HTC 02-JAN-2003  
BC042687  
BC042687.1 GI:27469363  
HTC  
Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg, R.  
1 (bases 1 to 2147)  
Direct Submission  
Submitted (02-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)  
Günaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsges, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Place: 6 Row: p Column: 3  
This clone has the following problem: no 5' EST match.

## FEATURES

```
Location/Qualifiers
1. .2147
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3157049"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
```

## ORIGIN

Query Match	66.1%;	Score 1438;	DB 11;	Length 2147;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1438;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	505	AGGAGATCTACTTCCCTCAACACCAATAAATGCCCTTCGCGGGGTGGCCAGTACACCGACA	564	
DB	1	AGGAGATCTACTTCCCTCAACACCAATAAATGCCCTTCGCGGGGTGGCCAGTACACCGACA	60	
QY	565	AGTTGTTCAAGTTCGCGCAATAAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG	624	
DB	61	AGTTGTTCAAGTTCGCGCAATAAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG	120	
QY	625	CCCGTGTGTGGCGAGCCTCTTTGCCGGACGCTCTGTGGCCGTGTGTGGACAGAAAGGCT	684	
DB	121	CCCGTGTGTGGCGAGCCTCTTTGCCGGACGCTCTGTGGCCGTGTGTGGACAGAAAGGCT	180	
QY	685	CTGGAGCGTACTCTATCTACATTTGCCAAATPACGCCCTACGGTAATGTGGGCCCTGTATGCC	744	
DB	181	CTGGAGCGTACTCTATCTACATTTGCCAAATPACGCCCTACGGTAATGTGGGCCCTGTATGCC	240	
QY	745	TCATTGAAATGGACCCCTGAGGCCAGTGACCTCTCCGGGGGCATTTCTGGCGCTCAGAGATG	804	
DB	241	TCATTGAAATGGACCCCTGAGGCCAGTGACCTCTCCGGGGGCATTTCTGGCGCTCAGAGATG	300	
QY	805	TGGCTGTGTAGGCTTGGGGTCAGCAATATACAGGGGGCCGAGCGCTCAGGTGGGGCCCCA	864	
DB	301	TGGCTGTGTAGGCTTGGGGTCAGCAATATACAGGGGGCCGAGCGCTCAGGTGGGGCCCCA	360	
QY	865	TCCTCAGCAGCGTCCCTCGGATATCTTCGACCAATAGAAATGGGCCCTAACTTCCTTT	924	
DB	361	TCCTCAGCAGCGTCCCTCGGATATCTTCGACCAATAGAAATGGGCCCTAACTTCCTTT	420	
QY	925	TCCACAAACCGGGCGATGGCACCTTTTGTGACCGTTCGGCCAGTGTGTGTGGACGACC	984	
DB	421	TCCACAAACCGGGCGATGGCACCTTTTGTGACCGTTCGGCCAGTGTGTGTGGACGACC	480	
QY	985	CCCAACCAAGATGGGCGAGGTGTGCCCTTGGCTGACTTCAACCGTGTATGGCAAAAGTGGACA	1044	
DB	481	CCCAACCAAGATGGGCGAGGTGTGCCCTTGGCTGACTTCAACCGTGTATGGCAAAAGTGGACA	540	
QY	1045	TCGTCATGGCAATGGAAATGGCCCCCACCGCCTCTATCTGCAAAATGAGCACCCATGGGA	1104	
DB	541	TCGTCATGGCAATGGAAATGGCCCCCACCGCCTCTATCTGCAAAATGAGCACCCATGGGA	600	
QY	1105	AGGTCCCGTTCGGGACATCGCCTCAACCCAAAGTTCTCCATGCCCTCCCTGTTCGGACGG	1164	
DB	601	AGGTCCCGTTCGGGACATCGCCTCAACCCAAAGTTCTCCATGCCCTCCCTGTTCGGACGG	660	
QY	1165	TCATCACCGCGCATTTTGCAATGACAGGAGCTGGAGATCTTCTTCAACCAATTCGCT	1224	

661	Db		TCATCACCGCGCACTTTGACAAATGACCCAGGAGCTGGAGATCTCTTCAACAACATTGGCT	720
1225	Qy		ACCGCAGCTCCTCAGCCAACCGCTCTTCGGCGTCATCCGTAGAGACACGGAGACCCCC	1284
721	Db		ACCGCAGCTCCTCAGCCAACCGCTCTTCGGCGTCATCCGTAGAGACACGGAGACCCCC	780
1285	Qy		TCATCAGGAGAGTCAATCCCGCGACGCTTGGAGCCTGAGGGCCGGGGCACAGGGGTG	1344
781	Db		TCATCAGGAGAGTCAATCCCGCGACGCTTGGAGCCTGAGGGCCGGGGCACAGGGGTG	840
1345	Qy		TGCTGACCGACTTCGACGGAGACGGGATGCTGAACTCATCTGTGCCATGGAGAGTCCA	1404
841	Db		TGCTGACCGACTTCGACGGAGACGGGATGCTGAACTCATCTGTGCCATGGAGAGTCCA	900
1405	Qy		TGGCTCAGCGCGTGTCCGCTTCCCGGGCAATCAGGGCTTCAACAACAACCTGGCTGCGAG	1464
901	Db		TGGCTCAGCGCGTGTCCGCTTCCCGGGCAATCAGGGCTTCAACAACAACCTGGCTGCGAG	960
1465	Qy		TGCTGCCACGCACCCGGTTTGGCGCTTTCGCCAGGGGAGCTAAGTCGTGCTCTACACCA	1524
961	Db		TGCTGCCACGCACCCGGTTTGGCGCTTTCGCCAGGGGAGCTAAGTCGTGCTCTACACCA	1020
1525	Qy		AGAAGAGTGGGGCCCACTGAGAGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGG	1584
1021	Db		AGAAGAGTGGGGCCCACTGAGAGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGG	1080
1585	Qy		AGCCGCTGGCACACTTTGGCTGGGAAAGATGAAGCCAGCAGTGTGGAGGTGACCTGTC	1644
1081	Db		AGCCGCTGGCACACTTTGGCTGGGAAAGATGAAGCCAGCAGTGTGGAGGTGACCTGTC	1140
1645	Qy		CAGATGCCAAGATCGTGAAGCCGAAAGTGCCACAGCGGGAGATGAATCAGTCTCTGGAGA	1704
1141	Db		CAGATGCCAAGATCGTGAAGCCGAAAGTGCCACAGCGGGAGATGAATCAGTCTCTGGAGA	1200
1705	Qy		TCCTCTACCCCGGGATGAGGACACACTTCAGAACCCAGCCCACTGGAGTGTGGCCAAAG	1764
1201	Db		TCCTCTACCCCGGGATGAGGACACACTTCAGAACCCAGCCCACTGGAGTGTGGCCAAAG	1260
1765	Qy		GATTCTCCAGCAGGAANAATGGCCATGTGATGCACCAATGAATCATCCAGTCTCCCAT	1824
1261	Db		GATTCTCCAGCAGGAANAATGGCCATGTGATGCACCAATGAATCATCCAGTCTCCCAT	1320
1825	Qy		TCGTGTGCCCTCGAGACAAGCCCGTATGTCTCAACACTATGGAAGCTACAGTGTCCGGA	1884
1321	Db		TCGTGTGCCCTCGAGACAAGCCCGTATGTCTCAACACTATGGAAGCTACAGTGTCCGGA	1380
1885	Qy		CCAAACAAGAGTGCACTGGGGCTACAGAGCCCAACGAGATGGCAAGCCTGCGTGGG	1942
1381	Db		CCAAACAAGAGTGCACTGGGGCTACAGAGCCCAACGAGATGGCAAGCCTGCGTGGG	1438

RESULT 2  
BT550776

BI550776	781 bp	mRNA	linear	EST 05-SEP-2001			
LOCUS	603195670F1 NIH_MGC_95	Homo sapiens	cDNA clone IMAGE:5275168 5',				
DEFINITION	mRNA sequence.						
ACCESSION	BI550776						
VERSION	BI550776.1	GI:15438088					
KEYWORDS	EST.						
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
REFERENCE	1 (bases 1 to 781)						
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .						
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)						
JOURNAL	Unpublished (1999)						
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapsb@mai.nih.gov">cgapsb@mai.nih.gov</a> Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)						

cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11694 row: 1 column: 17  
High quality sequence stop: 763.

## FEATURES

source  
1. .781  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5275168"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_95"  
/notes="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
(gtcgag); Oligo-dT primed using primer  
5'-TTTATTTTTATTTTATVN-3', size-selected for average  
insert size 2.5 kb and normalized to R0T 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: this  
is a NIH\_MGC Library."

ORIGIN  
Query Match 33.0%; Score 719; DB 12; Length 781;  
Best Local Similarity 99.9%; Pred. No. 5e-240;  
Matches 769; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGAGGCTCAGGCGACCGCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGCGG 62  
Db  
QY 4 GGAGGCTCAGGCGACCGCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGG 63  
Db  
QY 63 AGAGGCGGACGAGCGCGCGCTTCCACGCCCTAGGCGGCGGCGGCGGAGG 122  
Db  
QY 64 AGAGGCGGACGAGCGCGCGCTTCCACGCCCTAGGCGGCGGCGGAGG 123  
Db  
QY 123 ATGGCTCCGAGCGCTGACCCCGGCGATGTCAGGATGTTACCGGTTCTGCTGCTGCTG 182  
Db  
QY 183 TTTCTGCCATCAGTACGAGGGTCCAGCGGGTGAAACCCATGTTACTGAGTCAACCAAC 242  
Db  
QY 184 TTTCTGCCATCAGTACGAGGGTCCAGCGGGTGAAACCCATGTTACTGAGTCAACCAAC 243  
Db  
QY 243 TCAGTTCTGCTCCTGACTATGACAGTATCCACCCAGCTCAACTATGTTGTCAGTT 302  
Db  
QY 244 TCAGTTCTGCTCCTGACTATGACAGTATCCACCCAGCTCAACTATGTTGTCAGTT 303  
Db  
QY 303 ACTGATGAGCATGATGGGACCTTTGAGATCGTGTGGCGGGTACAAATGACCCCAAC 362  
Db  
QY 304 ACTGATGAGCATGATGGGACCTTTGAGATCGTGTGGCGGGTACAAATGACCCCAAC 363  
Db  
QY 363 CTGTTCTGAAGTATGACCGGCGCCCGAGCGGCTGTTGAAATCGCGGTGATGAGCGC 422  
Db  
QY 364 CTGTTCTGAAGTATGACCGGCGCCCGAGAGCGGCTGTTGAAATCGCGGTGATGAGCGC 423  
Db  
QY 423 AGCTCACCCCTACTAGCGCTGCGGACCGGCGGGAACGCCATTTGGGTTCACGCTTC 482  
Db  
QY 424 AGCTCACCCCTACTAGCGCTGCGGACCGGCGGGAACGCCATTTGGGTTCACGCTTC 483  
Db  
QY 483 GACATCGAGCGGACCGCGGAGGAGATCTACTTCTCAACACCAATATGCTTCTCG 542  
Db  
QY 484 GACATCGAGCGGACCGCGGAGGAGATCTACTTCTCAACACCAATATGCTTCTCG 543  
Db  
QY 543 GGGGTGGCCACGTACACGCAAGTTGTTCAAGTTCCGCAATACCGGTGGGAAGACATC 602  
Db  
QY 544 GGGGTGGCCACGTACACGCAAGTTGTTCAAGTTCCGCAATACCGGTGGGAAGACATC 603  
Db  
QY 603 CTGAGCGATGAGGTCAACGTGCGCGCTGTGTGGCCAGCGCTTTTGGCGGACGCTCTGTG 662

Db 604 CTGAGCGATGAGGTCAACGTGGCGCGTGTGTGGCCAGCGCTTTTGGCGGACGCTCTGTG 663  
QY 663 GCCTGTGTGACAGAAAGGGCTCTGGAGCGTACTCTATCTACATTCGCAATTAGCGCTAC 722  
Db 664 GCCTGTGTGACAGAAAGGGCTCTGGAGCGTACTCTATCTACATTCGCAATTAGCGCTAC 723  
QY 723 GGTAAATGTGGCGCTGATGCCCTCAATGAATGACCGCTGAGGCGCAGTGA 772  
Db 724 GGTAAATGTGGCGCTGATGCCCTCAATGAATGACCGCTGAGGCGCAGTGA 773

## RESULT 3

BG708845

LOCUS

DEFINITION

BG708845

VERSION

SOURCE

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG708845 841 bp mRNA linear EST 07-MAY-2001  
602673155F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4795957 5',  
mRNA sequence.

BG708845 GI:13986590  
EST.  
Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 841)  
NIH-MGC http://mgs.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLAM10679 row: m column: 14  
High quality sequence stop: 822.  
Location/Qualifiers  
1. .841

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4795957"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_96"

/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
(gtcgag); Oligo-dT primed using primer  
5'-TTTATTTTTATTTTATVN-3', size-selected for average  
insert size 2.3 kb and normalized to R0T 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIH/NHGRI, National Institutes of Health). Note: this is  
a NIH\_MGC Library."

## FEATURES

source

## ORIGIN

Query Match 30.0%; Score 653; DB 12; Length 841;  
Best Local Similarity 99.9%; Pred. No. 4.6e-217;  
Matches 703; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGAGGCTCAGGCGACCGCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGGCGG 62

Db 4 GGAGGCTCAGGCGACCGCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGGCGG 63

QY 63 AGAGCGGACGAGCGCGCGCTTCCACGCCCTAGGCGGCGGCGGCGGAGCGGAGG 122

Db 64 AGAGCGGACGAGCGCGCGCTTCCACGCCCTAGGCGGCGGCGGCGGAGCGGAGG 123



```
RESULT 5
BF727258      636 bp      mRNA      linear      EST 05-JAN-2001
LOCUS      by19a09.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
DEFINITION      sapiens cDNA clone by19a09 5', mRNA sequence.
ACCESSION      BF727258
VERSION      BF727258.1 GI:12043169
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      1 (bases 1 to 636)
JOURNAL      Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
COMMENT      NEIBANK: EST analysis and bioinformatics for ocular genomics
              Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
              Contact: Wistow G
              Section on Molecular Structure and Function
              National Eye Institute
              6/331, NIH, Bethesda, MD 20892-2740, USA
              Tel: 301 402 3452
              Fax: 301 496 0078
              Email: graeme@helix.nih.gov
              Plate: 19 row: a column: 09
              Seq primer: M13RPI reverse primer (ABI).
              Location/Qualifiers
                1..636
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="by19a09"
                  /tissue_type="lens"
                  /dev_stage="Adult"
                  /lab_host="EMDH10B"
                  /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                  BY"
                  /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                  from different adults (both approximately 40 years old)
                  together yielded 20ug of total RNA and 150ng mRNA for cDNA
                  library synthesis. A directionally cloned cDNA library in
                  the pCMVSPORT6 vector was constructed at Life
                  Technologies, essentially following the protocols of the
                  SuperScript Plasmid System full details of which are
                  contained in the manufacturer's instruction manual
                  (http://www.lifetech.com/). First strand synthesis was
                  carried out using a Not I primer-adaptor
                  [5'-pGACTAGTCTAGATCGGCGCGCC(T)15-3']. Not I/blunt
                  end inserts were cloned into the Not I/EcoR V sites in the
                  vector. EST analysis was performed on the unamplified
                  library at the NIH Intramural Sequencing Center (NISC)."
```

Query Match 28.8%; Score 626; DB 10; Length 636;  
Best Local Similarity 100.0%; Pred. No. 1.3e-207;  
Matches 626; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1277 AGACCCCTCATCGAGAGCTCAATCCGCGCAGCGCTTGGAGCCTCAGCGCGGCGAC 1336
DB 11 AGACCCCTCATCGAGAGCTCAATCCGCGCAGCGCTTGGAGCCTCAGCGCGGCGAC 70
QY 1337 AGGGGGTGTGTCACCGACTTCGACGAGACGGGATCTGGACCTCATCTTGTCCCATGG 1396
DB 71 AGGGGGTGTGTCACCGACTTCGACGAGACGGGATCTGGACCTCATCTTGTCCCATGG 130
QY 1397 AGAGTCCATCGGCTCAGCGCTGTCCCTCTTCCGGGGCAATCAGGGCTTCAACAACAACATG 1456
DB 131 AGAGTCCATCGGCTCAGCGCTGTCCCTCTTCCGGGGCAATCAGGGCTTCAACAACAACATG 190
QY 1457 GCTCGAGTGTGTCACCGACCGCGTTTGGGGCTTTGCCAGGGGAGCTAAGGTCTGTGT 1516
DB 191 GCTCGAGTGTGTCACCGACCGCGTTTGGGGCTTTGCCAGGGGAGCTAAGGTCTGTGT 250
QY 1517 CTACACCAAGAGATGGGGGCCCACTTGAGGATCATCGACGGGGGCTCAGGCTACCTGTG 1576
```

```
DB 251 CTACACCAAGAGATGGGGGCCCACTTGAGGATCATCGACGGGGCTCAGGCTACCTGTG 310
QY 1577 TGAGATGGAGCCCGTGGCACACATTTGSCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGT 1636
DB 311 TGAGATGGAGCCCGTGGCACACATTTGSCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGT 370
QY 1637 GACGTGGCCAGATGGCAAGATGCTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGT 1696
DB 371 GACGTGGCCAGATGGCAAGATGCTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGT 430
QY 1697 GCTGGAGATCTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCATCTGGAGTG 1756
DB 431 GCTGGAGATCTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCATCTGGAGTG 490
QY 1757 TGCCCAAGGATTTCTCCAGCAGGAAATGCCATTGCATGACCAATGAATGCATCCA 1816
DB 491 TGCCCAAGGATTTCTCCAGCAGGAAATGCCATTGCATGACCAATGAATGCATCCA 550
QY 1817 GTTCCCATTTGTTGCCCTCGACACAAGCCGCTATGTGTCAACACCTATGGAAGCTACAG 1876
DB 551 GTTCCCATTTGTTGCCCTCGACACAAGCCGCTATGTGTCAACACCTATGGAAGCTACAG 610
QY 1877 GTCCCGGACCAACAAGAGTGCAGTC 1902
DB 611 GTCCCGGACCAACAAGAGTGCAGTC 636
```

RESULT 6  
LOCUS BQ067249 1101 bp mRNA linear EST 02-APR-2002  
DEFINITION AGENCOURT\_6767325 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5751633  
5', mRNA sequence.  
ACCESSION BQ067249  
VERSION BQ067249.1 GI:19896295  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 1101)  
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM12784 row: i column: 10  
 High quality sequence stop: 612.  
 Location/Qualifiers  

FEATURES  
source

## ORIGIN

Query Match 28.2%; Score 613; DB 13; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 3.4e-203;  
Matches 613; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 720 TACGGTAATGGGCGCTGATGCCCTCATTTGAATGGACCTTGAGCCAGTGACCTCTCC 779  
DB 1 TACGGTAATGGGCGCTGATGCCCTCATTTGAATGGACCTTGAGCCAGTGACCTCTCC 60

QY 780 CGGGGCAATTCGGGCTCAGAGATGTGCTGTGAGGCTGGGGTCAGAAATATACAGGG 839  
DB 61 CGGGGCAATTCGGGCTCAGAGATGTGCTGTGAGGCTGGGGTCAGAAATATACAGGG 120

QY 840 GCGGAGCGCTCAGGCTGGGCGCCATCTCAGCAGCAGTGTGCTCGATATCTTCTGGAC 899  
DB 121 GCGGAGCGCTCAGGCTGGGCGCCATCTCAGCAGCAGTGTGCTCGATATCTTCTGGAC 180

QY 900 AATGAGATGGGCTAACTTCTTTTCCAAACCGGGCGATGGCACCTTTGTGGACGCT 959  
DB 181 AATGAGATGGGCTAACTTCTTTTCCAAACCGGGCGATGGCACCTTTGTGGACGCT 240

QY 960 GCGGCACTGTGTGTGAGAGAGACCCCAACAGCAGTGGGCGAGGTGTGCGCTGTGAC 1019  
DB 241 GCGGCACTGTGTGTGAGAGAGACCCCAACAGCAGTGGGCGAGGTGTGCGCTGTGAC 300

QY 1020 TTCAACCTGTATGCAAGTGGACATCTGTATGCACTGGAATGGCCCGCCAGCGCTC 1079  
DB 301 TTCAACCTGTATGCAAGTGGACATCTGTATGCACTGGAATGGCCCGCCAGCGCTC 360

QY 1080 TATCTGCAAAATGAGCACCCTCGGAAGTTCGGGAGCATCGCCTCACCCAAAGTTC 1139  
DB 361 TATCTGCAAAATGAGCACCCTCGGAAGTTCGGGAGCATCGCCTCACCCAAAGTTC 420

QY 1140 TCCATGCCCTCCCTGTCCGACCGGTATACCGCGCATTTTGAACAATGACAGAGCTG 1199  
DB 421 TCCATGCCCTCCCTGTCCGACCGGTATACCGCGCATTTTGAACAATGACAGAGCTG 480

QY 1200 GAGATCTTTTCACAAATTCCTACCGCAGCTCTCAGCAACCGCTCTTCCGCTC 1259  
DB 481 GAGATCTTTTCACAAATTCCTACCGCAGCTCTCAGCAACCGCTCTTCCGCTC 540

QY 1260 ATCCGTAGAGCAGCAGACCCCTCATCGAGGAGCTCAATCCCGGCGACCTTGGAG 1319  
DB 541 ATCCGTAGAGCAGCAGACCCCTCATCGAGGAGCTCAATCCCGGCGACCTTGGAG 600

QY 1320 CTGAGGCGCGGG 1332  
DB 601 CTGAGGCGCGGG 613

## RESULT 7

BI758703  
LOCUS 603024078F1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5194530 5',  
DEFINITION mRNA sequence.

ACCESSION BI758703  
VERSION BI758703.1 GI:15750294  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 770)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LAM11486 row: 1 column: 19

High quality sequence start: 3

High quality sequence stop: 770.

## FEATURES

Location/Qualifiers  
source  
1..770  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5194530"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_114"  
/note="Foram: brain; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains, age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH\_MGC Library."

## ORIGIN

Query Match 27.1%; Score 590; DB 12; Length 770;  
Best Local Similarity 99.8%; Pred. No. 4.1e-195;  
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCAGAGATGTTACCGTTCCTGCTGCTCT 180  
DB 109 GGATGGCTCCGAGCGCTGACCCCGGCATGTCAGAGATGTTACCGTTCCTGCTGCTCT 168

QY 181 GGTTCCTGCCCATCTCATGAGGGTCCCGAGCGGCTGAAACCCATGTTCACTGCAGTACCA 240  
DB 169 GGTTCCTGCCCATCTCATGAGGGTCCCGAGCGGCTGAAACCCATGTTCACTGCAGTACCA 228

QY 241 ACTCAGTTCCTGCTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGCAG 300  
DB 229 ACTCAGTTCCTGCTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGCAG 288

QY 301 TTACTGATGTGGACCATGATGGGATCTTGATGATCGTGGCGGGGTACATGGACCCA 360  
DB 289 TTACTGATGTGGACCATGATGGGATCTTGATGATCGTGGCGGGGTACATGGACCCA 348

QY 361 ACCTGGTTCGAGTATGACCGGGCCAGAGCGCTGTGAACATCGCGTTCGATGAGC 420  
DB 349 ACCTGGTTCGAGTATGACCGGGCCAGAGCGCTGTGAACATCGCGTTCGATGAGC 408

QY 421 GCAGCTCACCCCTACTACCGGTTCGGGACCGGAGGAAACCCCATTTGGGGTCAAGCCT 480  
DB 409 GCAGCTCACCCCTACTACCGGTTCGGGACCGGAGGAAACCCCATTTGGGGTCAAGCCT 468

QY 481 GCGCATCGACGGGAGCGCGGGAGGAGATCTTCTTCAACACCATTAATGCTTCT 540  
DB 469 GCGCATCGACGGGAGCGCGGGAGGAGATCTTCTTCAACACCATTAATGCTTCT 528

QY 541 CGGGGGTGGCCACGTACACCGCAAGTGTTCAGTTCGCAATAACCGGTGGGAAGACA 600  
DB 529 CGGGGGTGGCCACGTACACCGCAAGTGTTCAGTTCGCAATAACCGGTGGGAAGACA 588

QY 601 TCTGAGCGATGAGGTCAACGTTGGCCCGTGTGTGGCAGCGCTCTTTCGCGACGCTCTG 660  
DB 589 TCTGAGCGATGAGGTCAACGTTGGCCCGTGTGTGGCAGCGCTCTTTCGCGACGCTCTG 648

QY 661 TGGCCCTGTGTGACAGAAAGCGCTCTGACCGCTACTCTATCTACATTCGCAATACGCT 720  
DB 649 TGGCCCTGTGTGACAGAAAGCGCTCTGACCGCTACTCTATCTACATTCGCAATACGCT 708

QY 721 ACGGTAAATGTGGGCCCTGATGCCCTCAATTGAATGGACCT 761  
DB 709 ACGGTAAATGTGGGCCCTGATGCCCTCAATTGAATGGACCT 749

```

RESULT 8
BF527664          926 bp      mRNA      linear      EST 11-DEC-2000
LOCUS             602040644F1 NCI CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4178363
DEFINITION        5', mRNA sequence.
ACCESSION         BF527664
VERSION           BF527664.1 GI:11615027
KEYWORDS          EST.
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 926)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs-r@mail.nih.gov
                  Tissue Procurement: David N. Louis, M.D.
                  CDNA Library Preparation: Life Technologies, Inc.
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LHAM9486 row: h column: 12
                  High quality sequence stop: 691.
                  Location/Qualifiers
                    1..926
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4178363"
                      /tissue_type="anaplastic oligodendroglioma with 1p/19q
                      loss"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /clone_lib="NCI_CGAP_Brn67"
                      /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
                      Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                      Average insert size 2.3 kb. Constructed by Life
                      Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source
    Query Match      26.0%; Score 566; DB 10; Length 926;
    Best Local Similarity 100.0%; Pred. No. 8.5e-187;
    Matches 566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 AGGAGATCTACTTCTCAACACCAACCAATAGCCCTTCGCGGGGTGGCCAGTACACCGACA 564
DB      |||||||
DB      1 AGGAGATCTACTTCTCAACACCAACCAATAGCCCTTCGCGGGGTGGCCAGTACACCGACA 60

QY 565 AGTTGTTCAAGTTCCGCAATACCGGTGGGAAGACATCTGAGCGATGAGGTCAACGTGG 624
DB      |||||||
DB      61 AGTTGTTCAAGTTCCGCAATACCGGTGGGAAGACATCTGAGCGATGAGGTCAACGTGG 120

QY 625 CCGCTGGTGTGGCGAGCCTTTTGGCGACGCTCTGTGTCGCTGTGTGGCAGAAAGGGCT 684
DB      |||||||
DB      121 CCGCTGGTGTGGCGAGCCTTTTGGCGACGCTCTGTGTCGCTGTGTGGCAGAAAGGGCT 180

QY 685 CTGGAGCGTACTCTACTACATTGCCAATTACGCTTACGCTATGTGGCCCTGTATGCC 744
DB      |||||||
DB      181 CTGGAGCGTACTCTACTACATTGCCAATTACGCTTACGCTATGTGGCCCTGTATGCC 240

QY 745 TCATTGAATGGACCTTGAGGCGAGTGACCTCTCCCGGGGCAATCTGGCGCTCAGAGATG 804
DB      |||||||
DB      241 TCATTGAATGGACCTTGAGGCGAGTGACCTCTCCCGGGGCAATCTGGCGCTCAGAGATG 300

QY 805 TGGCTGCTGAGGCTGGGGTTCAGAAATATACAGGGGGCCGAGGCGCTCAGCGTGGGCCCA 864
DB      |||||||
DB      301 TGGCTGCTGAGGCTGGGGTTCAGAAATATACAGGGGGCCGAGGCGCTCAGCGTGGGCCCA 360

QY 865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGCAATGAGAAATGGGCGCTAATCTTCCTTT 924

```

```

DB      |||||||
DB      361 TCCTCAGCAGCAGTGCCTCGGATATCTTCGCAATGAGTGGGCTAATCTTCCTTT 420

QY 925 TCCACAAACCGGGGCGATGGCACCTTTTGTGACGCTGGCGCCAGTCTGCTGGTGGACGACC 984
DB      |||||||
DB      421 TCCACAAACCGGGGCGATGGCACCTTTTGTGACGCTGGCGCCAGTCTGCTGGTGGACGACC 480

QY 985 CCCACCAAGCATGGCGAGGTGTCCCTCGCTGCTGCTCAACCCGATGGCAAGTGGACA 1044
DB      |||||||
DB      481 CCCACCAAGCATGGCGAGGTGTCCCTCGCTGCTGCTCAACCCGATGGCAAGTGGACA 540

QY 1045 TCGTCTATGGCACTGGAATGGCCCC 1070
DB      |||||||
DB      541 TCGTCTATGGCACTGGAATGGCCCC 566

RESULT 9
BX451456          904 bp      mRNA      linear      EST 22-MAY-2003
LOCUS             CS0DF014VH11 5-PRIME, mRNA sequence.
DEFINITION        BX451456 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
ACCESSION         BX451456
VERSION           BX451456.1 GI:31022247
KEYWORDS          EST.
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 904)
AUTHORS           Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE             Full-length cDNA libraries and normalization
JOURNAL           Unpublished (2001)
COMMENT           Contact: Genoscope
                  Genoscope Centre National de Sequencage
                  BP 191 91006 EVRY cedex - France
                  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
                  Library was constructed by Life Technologies, a division of
                  Invitrogen. This sequence belongs to sequence cluster 5247.r For
                  more information about this cluster, see
                  http://www.genoscope.cns.fr/
                  cgi-bin/cluster.cgi?seq=CS0BAF012ZE12_AF01150_1&cluster=5247.r.
                  Contact : Feng Liang Email : fliang@lifetech.com URL :
                  http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
                  Faraday Avenue Genoscope sequence ID : CS0BAF012ZE12_AF01150_1.
                  Location/Qualifiers
                    1..904
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DF014VH11"
                      /tissue_type="FETAL BRAIN"
                      /dev_stage="fetal"
                      /clone_lib="Homo sapiens FETAL BRAIN"
                      /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                      was primed with a NotI-oligo (dT) primer. Five prime end
                      enriched, double-strand cDNA was digested with Not I and
                      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                      vector. Library was not normalized."

FEATURES
source
    Query Match      25.4%; Score 553; DB 13; Length 904;
    Best Local Similarity 100.0%; Pred. No. 2.9e-182;
    Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 CAAACCGGGCGATGGCACCTTTTGTGACGCTGGCCAGTCTGCTGGTGGACGACCCCA 988
DB      |||||||
DB      90 CAAACCGGGCGATGGCACCTTTTGTGACGCTGGCCAGTCTGCTGGTGGACGACCCCA 149

QY 989 CCAGCATGGCGAGGTGTGGCCCTGGCTGACTTCAACCGTGTATGCAAGTGGACATCGT 1048
DB      |||||||
DB      150 CCAGCATGGCGAGGTGTGGCCCTGGCTGACTTCAACCGTGTATGCAAGTGGACATCGT 209

QY 1049 CTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAATGAGCACCCACCGAAGT 1108

```



```
Db 210 CTATGGCACTGGAAATGGCCCCCAGCGCTCTATCTGCAATGAGCACCCTATGGAGGT 269
Qy 1109 CCGCTTCGGGACATCGCTCACCAAGTTCTCCATGCCCTCCCTGTCGCCGACGGTCAAT 1168
Db 270 CCGCTTCGGGACATCGCTCACCAAGTTCTCCATGCCCTCCCTGTCGCCGACGGTCAAT 329
Qy 1169 CACGCCGACATTGACAAATGACAGGAGCTGGAGATCTTCTTCAACACATGCTACCG 1228
Db 330 CACGCCGACATTGACAAATGACAGGAGCTGGAGATCTTCTTCAACACATGCTACCG 389
Qy 1229 CAGCTCTCCAGCAACCCGCTCTCCGCGTCATCGGTAGAGACGACGAGACCCCTCAT 1288
Db 390 CAGCTCTCCAGCAACCCGCTCTCCGCGTCATCGGTAGAGACGACGAGACCCCTCAT 449
Qy 1289 CGAGAGCTCAATCCCGCGACGCTTCGAGCCTGAGCGCGCGGCGCACAGGGGGTGTGT 1348
Db 450 CGAGAGCTCAATCCCGCGACGCTTCGAGCCTGAGCGCGCGGCGCACAGGGGGTGTGT 509
Qy 1349 GACCGACTTCAGCGAGAGCGGATCTGGACCTCATCTTGTCCCATGGAGAGTCCATGCG 1408
Db 510 GACCGACTTCAGCGAGAGCGGATCTGGACCTCATCTTGTCCCATGGAGAGTCCATGCG 569
Qy 1409 TCAGCGCTGTCCTGCTTCCTCCGGGCAATCAGGGCTTCAACACACTGCTGCGAGTGT 1468
Db 570 TCAGCGCTGTCCTGCTTCCTCCGGGCAATCAGGGCTTCAACACACTGCTGCGAGTGT 629
Qy 1469 GCCACGACCCCG 1481
Db 630 GCCACGACCCCG 642
```

```
RESULT 10
BG911024
LOCUS
DEFINITION 602813023F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4944964
5', mRNA sequence.
```

```
ACCESSION BG911024.1 GI:14291500
VERSION BG911024
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLAM10890 row: n column: 05
High quality sequence stop: 656.
Location/Qualifiers
```

```
FEATURES
source
1..735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4944964"
/tissue_type="anaplastic oligodendroglioma with ip/19q
loss"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Brn67"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: Not I;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
```

## ORIGIN

```
Query Match 23.8%; Score 519; DB 12; Length 735;
Best Local Similarity 99.8%; Pred. No. 2.2e-170;
Matches 569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 199 AGGGGTCCCAGCGGCTGAACCCATGTTCACTGAGTCAACCACTCAGTTCTGCTCTCTG 258
Db 1 AGGGGTCCCAGCGGCTGAACCCATGTTCACTGAGTCAACCACTCAGTTCTGCTCTCTG 60
Qy 259 ACTATGACAGTAATCCCAACCCAGCTCAACTATGTTGTCGAGTTACTCATGTGGACCATG 318
Db 61 ACTATGACAGTAATCCCAACCCAGCTCAACTATGTTGTCGAGTTACTCATGTGGACCATG 120
Qy 319 ATGGGACCTTTGAGATCGTCTGGCGGGTCAATGAGACCCCAACCTGTTCTGAAGTATG 378
Db 121 ATGGGACCTTTGAGATCGTCTGGCGGGTCAATGAGACCCCAACCTGTTCTGAAGTATG 180
Qy 379 ACCGGGCCCCAGAGCGGCTGTGAACATCGCGGTGATGAGGAGCTCACCCTACTACG 438
Db 181 ACCGGGCCCCAGAGCGGCTGTGAACATCGCGGTGATGAGGAGCTCACCCTACTACG 240
Qy 439 CGCTCGGGACCGGACGGGAGCGCATTTGGGGTCAACAGCTTCGACATCGACGGGAGC 498
Db 241 CGCTCGGGACCGGACGGGAGCGCATTTGGGGTCAACAGCTTCGACATCGACGGGAGC 300
Qy 499 GCGGGAGGAGATCTACTTCTCTCAACACCAATTAATGCTTCTCGGGGGTGGCCAGTACA 558
Db 301 GCGGGAGGAGATCTACTTCTCTCAACACCAATTAATGCTTCTCGGGGGTGGCCAGTACA 360
Qy 559 CCGACAAAGTTTCAAGTTCGCAATAACCGTGGGAGACATCTGAGCGATGAGGTCA 618
Db 361 CCGACAAAGTTTCAAGTTCGCAATAACCGTGGGAGACATCTGAGCGATGAGGTCA 420
Qy 619 AGTGGCGCGGTGTGGCCAGCTCTTTGCCGACGCTCTGTGGCGCTGTGTGGACAGAA 678
Db 421 AGTGGCGCGGTGTGGCCAGCTCTTTGCCGACGCTCTGTGGCGCTGTGTGGACAGAA 480
Qy 679 AGGGCTCTGGACGCTACTCTATCTATTCATTCGCAATTAACGCTACGCTACGTAATGTGGGCCCTG 738
Db 481 AGGGCTCTGGACGCTACTCTATCTATTCATTCGCAATTAACGCTACGCTACGTAATGTGGGCCCTG 540
Qy 739 ATGGCTCTATTTGAATGACCTCTGAGGCCA 768
Db 541 ATGGCTCTATTTGAATGACCTCTGAGGCCA 570
```

## RESULT 11

```
BM707643
LOCUS
DEFINITION UI-B-C11-afs-0-03-0-UI.r1 UI-E-C11 Homo sapiens cDNA clone
UI-B-C11-afs-0-03-0-UI 5', mRNA sequence.
ACCESSION BM707643
VERSION BM707643.1 GI:19020901
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 567)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4155 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
```

Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 Location/Qualifiers  
 1..567  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-C11-afs-o-03-0-UI"  
 /tissue\_type="RE and Choroid"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /notes="UI-E-C11"  
 /notes="Organ: eye; Vector: pUT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-C11 is a normalized cDNA library containing the following tissue(s): RE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pUT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN  
 Query Match 23.7%; Score 516; DB 12; Length 567;  
 Best Local Similarity 99.8%; Pred. No. 2.7e-169;  
 Matches 566; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 880 CTTGGATATCTTCTGCAATGAGAATGGGCTAACTTCCTTTTCCAAACCGGGGCG 939  
 Db 1 CCTCGATATCTTCTGCAATGAGAATGGGCTAACTTCCTTTTCCAAACCGGGGCG 60  
 QY 940 ATGGACCTTTGGAGCGTGGCCAGTCTGCTGTGGACGACCCACCACGATGGGC 999  
 Db 61 ATGGACCTTTGGAGCGTGGCCAGTCTGCTGTGGACGACCCACCACGATGGGC 120  
 QY 1000 GAGGTGCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGACATCGTCTATGGCAACT 1059  
 Db 121 GAGGTGCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGACATCGTCTATGGCAACT 180  
 QY 1060 GGAAATGGCCCAACCGCTCTATCTGCAATGAGCAACCCATGGGAAGTCCGCTCCGGG 1119  
 Db 181 GGAAATGGCCCAACCGCTCTATCTGCAATGAGCAACCCATGGGAAGTCCGCTCCGGG 240  
 QY 1120 ACATCGCCTCACCCAGTTCTCCATGCCCTCCCTGTCGACGATCATCACCGCGACT 1179  
 Db 241 ACATCGCCTCACCCAGTTCTCCATGCCCTCCCTGTCGACGATCATCACCGCGACT 300  
 QY 1180 TTGACAAATGACGAGGCTGGAGATCTTCTTCAACCAATTTGCTACCGCAGTCTCTCAG 1239  
 Db 301 TTGACAAATGACGAGGCTGGAGATCTTCTTCAACCAATTTGCTACCGCAGTCTCTCAG 360  
 QY 1240 CCACCGCCTCTTTCGCGTCAATCCGTAGAGACACGAGACCCCTCATCGAGAGCTCA 1299  
 Db 361 CCACCGCCTCTTTCGCGTCAATCCGTAGAGACACGAGACCCCTCATCGAGAGCTCA 420  
 QY 1300 ATCCCGGACGCTTGGAGCCTGAGGCGGGGCAAGGGGGTGTGGTGAACGACTTCG 1359  
 Db 421 ATCCCGGACGCTTGGAGCCTGAGGCGGGGCAAGGGGGTGTGGTGAACGACTTCG 480  
 QY 1360 ACGGAGACGGGATGTGGACCTCATCTGTGCCATGGAGTCCATGGCTCAGCCCTGT 1419

Db 481 ACGGAGACGGGATGCTGGACCTCATCTTGTCATGGAGAGTCCATGCTCAGCGCTGT 540  
 QY 1420 CCGTCTTCCGGGGCAATCAGGGCTTCA 1446  
 Db 541 CCGTCTTCCGGGGCAATCAGGGCTTCA 567  
 RESULT 12  
 BI596960 1063 bp mRNA linear EST 07-SEP-2001  
 LOCUS 603242976F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5285419 5',  
 mRNA sequence.  
 ACCESSION BI596960  
 VERSION BI596960.1 GI:15489899  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1063)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1AM11721 row: g column: 20  
 High quality sequence stop: 756.  
 Location/Qualifiers  
 1..1063  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5285419"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH MGC 96"  
 /notes="Organ: brain; Vector: pBluescriptR (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
 (gtcgag); Oligo-dT primed using primer  
 5'-TTTTTTTTTTTTTNN-3', size-selected for average  
 insert size 2.3 kb and normalized to 10^5. This is a  
 primary library enriched for full-length clones and  
 constructed using the Cap-trapper method (Carninci, in  
 preparation). Library constructed by M. Brownstein  
 (NIH/NHGRI, National Institutes of Health). Note: this is  
 a NIH\_MGC Library."

ORIGIN  
 Query Match 23.6%; Score 514; DB 12; Length 1063;  
 Best Local Similarity 99.8%; Pred. No. 1e-168;  
 Matches 564; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 GGAGGCTCAGGCGACCGCGGCTCGGAGCAAGAGCGGCGCGCGCGG 62  
 Db 4 GGAGGCTCAGGCGACCGCGGCTCGGAGCAAGAGCGGCGCGCGCGG 63  
 QY 63 AGAGGCGGACGAGCGCGCGGCTTCCAGCCCTTAGCGGCGGCGCGGAGG 122  
 Db 64 AGAGGCGGACGAGCGCGCGGCTTCCAGCCCTTAGCGGCGGCGCGGAGG 123  
 QY 123 ATGGGCTCCGAGCGCTGACCCCGGCGATGCCAGGATGTACCGTCTCTGCTGCTCG 192  
 Db 124 ATGGGCTCCGAGCGCTGACCCCGGCGATGCCAGGATGTACCGTCTCTGCTGCTCG 183  
 QY 183 TTCTCTGCCATCACTAGGGGTCCCGGCGGTGAACCCATGTTTCACTCAGTCAACCAAC 242

184 TTTCTGCCCATCACTAGGGGTCCAGCGGCTGAACCCATGTTTCACTGCAGTCAACCAAC 243  
243 TCAGTTCTGCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGTCAGTT 302  
244 TCAGTTCTGCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGTCAGTT 303  
303 ACTGATGCGACCATGATGGGACTTTGAGATCGTGTGGGGGTACAAATGACCCCAAC 362  
304 ACTGATGCGACCATGATGGGACTTTGAGATCGTGTGGGGGTACAAATGACCCCAAC 363  
363 CTGTTCTGAAGTATGACCGGCGCCAGAGCGGCTGGTGAACATCGCGTTCGATGAGCGC 422  
364 CTGTTCTGAAGTATGACCGGCGCCAGAGCGGCTGGTGAACATCGCGTTCGATGAGCGC 423  
423 AGCTCACCTTACTAGCGCTGCGGACCGGCGGAGGAGCGCCATGTTGGGTACACGCTGC 482  
424 AGCTCACCTTACTAGCGCTGCGGACCGGCGGAGGAGCGCCATGTTGGGTACACGCTGC 483  
483 GACATCGAGCGGCGGCGGAGGAGATCTACTTCTCAACCAATTAATGCTTCTCG 542  
484 GACATCGAGCGGCGGCGGAGGAGATCTACTTCTCAACCAATTAATGCTTCTCG 543  
543 GGGGTGGCCACGTACACCGACAAGT 567  
544 GGGGTGGCCACGTACACCGACAAGT 568

RESULT 13  
CD676152  
LOCUS  
DEFINITION  
fs32g06.y1 Human Lens cDNA (Normalized): fs Homo sapiens cDNA clone  
fs32g06.5, mRNA sequence.  
CD676152.1 GI:32177883  
VERSION  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Wistow,G., Bernstein,S.L., Wyatt,M.K., Behal,A., Touchman,J.W.,  
Bouffard,G., Smith,D. and Peterson,K.  
Expressed sequence tag analysis of adult human lens for the NEIRank  
project: over 2000 non-redundant transcripts, novel genes and  
splice variants  
Mol. Vis. 8 (4), 171-184 (2002)  
22103463  
12107413  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 32 row: g column: 06  
Seq primer: M13Rpl reverse primer (ABI).  
Location/Qualifiers  
1..578  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fs32g06"  
/tissue\_type="lens"  
/dev\_stage="Adult"  
/lab\_host="EMD10B"  
/note="Organ: Eye; Vector: pCMVSPORT6; A human lens  
library (by) was normalized by self-subtraction. One  
portion of double stranded plasmid DNA representing the  
library was linearized by NotI. This NotI digested library  
was used as a template for biotinylated RNA synthesis

using SP6 RNA polymerase. Another portion of the double  
stranded plasmid library was converted to single-stranded  
circles in vitro using Gene II and Exonuclease III (Life  
Technologies). Single-stranded DNA (1 mg) was hybridized  
(Cot 500) with 41 mg of Bio-RNA and vector blocking  
oligonucleotides. The hybridized Bio-RNA/ss-circles were  
removed by streptavidin:phenol extraction. EST analysis  
was performed on the library at the NIH Intramural  
Sequencing Center (NISC)."

## ORIGIN

Query Match 23.3%; Score 507; DB 14; Length 578;  
Best Local Similarity 99.8%; Pred. No. 3.6e-166;  
Matches 557; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 12 AGCCAGACCCCGGACCCGGGCTGGGAGCAAGCAGCGCGCGCGGAGAGCGGC 71  
Db 1 AGCCAGACCCCGGACCCGGGCTGGGAGCAAGCAGCGCGCGCGGAGAGCGGC 60  
QY 72 AGCGAGCGCCCGCTTTCCACGCCCTAGGCGCGCGGCGCGAGCGGAGATGGCTCG 131  
Db 61 AGCGAGCGCCCGCTTTCCACGCCCTAGGCGCGCGGCGCGAGCGGAGATGGCTCG 120  
QY 132 AGCGTGACCCCGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTGCTGCTG 191  
Db 121 AGCGTGACCCCGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTGCTGCTG 180  
QY 192 ATCACTGAGGGTCCCGAGCGGCTCAACCCATGTTCACTGCAGTCAACCACTCAGTTCG 251  
Db 181 ATCACTGAGGGTCCCGAGCGGCTCAACCCATGTTCACTGCAGTCAACCACTCAGTTCG 240  
QY 252 CTCCTCTGACTATGACAGTAATCCACCCAGTCAACTATGGTGTGGCAGTTACTGATG 311  
Db 241 CTCCTCTGACTATGACAGTAATCCACCCAGTCAACTATGGTGTGGCAGTTACTGATG 300  
QY 312 GACCATGATGGGACTTTGAGATCGTGTGGCGGGGTACAAATGGACCACTGGTTCG 371  
Db 301 GACCATGATGGGACTTTGAGATCGTGTGGCGGGGTACAAATGGACCACTGGTTCG 360  
QY 372 AAGTATGACCGCGCCAGAAAGCGGCTGGTGAACATCGCGGTGATGAGCGAGCTCACCC 431  
Db 361 AAGTATGACCGCGCCAGAAAGCGGCTGGTGAACATCGCGGTGATGAGCGAGCTCACCC 420  
QY 432 TACTACGCGCTCGGAGACCGGAGGAAACGCAATGGGGTCAAGCCTTGACATCGAC 491  
Db 421 TACTACGCGCTCGGAGACCGGAGGAAACGCAATGGGGTCAAGCCTTGACATCGAC 480  
QY 492 GGGAGCGCGGAGGAGATCTACTTCTCAACCAATAATGCTTCTCGGGGGTGGCC 551  
Db 481 GGGAGCGCGGAGGAGATCTACTTCTCAACCAATAATGCTTCTCGGGGGTGGCC 540  
QY 552 AGGTACACCGACAGTTG 569  
Db 541 AGGTACACCGACAGTTG 558

## RESULT 14

## BI458314

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

BI458314 804 bp mRNA linear EST 21-AUG-2001  
603199270F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5278624 5',  
mRNA sequence.  
BI458314  
BI458314.1 GI:15248970  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 804)  
NIH-MGC <http://img.ncbi.nlm.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.



Db	274	TACGCCTACGGTAATGTGGCGCCCTGATGCCCTCATTTGAATGACCCCTGAGGCC	333
Qy	768	AGTGACCTCTCCCGGGGCATTCTTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGTCACG	827
Db	334	AGTGACCTCTCCCGGGGCATTCTTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGTCACG	393
Qy	828	AAATATACAGGGGGCCGAGGCGTTCAGCGTGGGCCCATCTTCAGCAGCAGTGCCTCGAT	887
Db	394	AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCATCTTCAGCAGCAGTGCCTCGAT	453
Qy	888	ATCTTTCGCAATGAGAAATGGGCTAACTCTCTTTTCCACAAACGGGGCGATGGCAC	947
Db	454	ATCTTTCGCAATGAGAAATGGGCTAACTCTCTTTTCCACAAACGGGGCGATGGCAC	513
Qy	948	TTTGTGACGCTGCGG	963
Db	514	TTTGTGACGCTGCGG	529

Search completed: August 10, 2004, 03:32:25  
Job time : 5093 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 3, 2004, 19:36:29 ; Search time 100 Seconds  
(without alignments)  
12302.118 Million cell updates/sec

Title: US-09-914-958B-35  
Perfect score: 4056  
Sequence: 1 cggaggctcgaggccagcc.....tctcttgaaaaa 2177

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-Q=/cn2.1/USPTO.spool -DEV=xlp  
-DB=A Geneseq 29Jan04 -OFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914958 @CGN\_1.1.B1 @runat\_03082004 161651.29663 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3335	82.2	626	3 AAB08864	AAB08864 Amino aci
2	3292	81.2	618	5 AAE18681	AAE18681 Human SCI
3	3190.5	78.7	653	5 AA018264	AA018264 Human cho
4	3190.5	78.7	653	5 AAE18679	AAE18679 Human SCI
5	3190.5	78.7	653	6 ABR48171	ABR48171 Human bla
6	3190.5	78.7	653	6 ABU56636	ABU56636 Lung can
7	3181.5	78.4	637	5 AAE18682	AAE18682 Human SCI
8	3179	78.4	612	5 AAE18682	AAE18682 Human SCI
9	3137	77.3	636	5 AA018267	AA018267 Human cho
10	2337	57.6	1597	6 ABR48172	ABR48172 Human bla

11	2337	57.6	1597	6	ABU56637	ABU56637 Lung can
12	2124.5	52.4	435	4	AA039983	AA039983 Human pol
13	1967.5	48.5	418	4	ABR48170	ABR48170 Human bla
14	1967.5	48.5	418	6	ABR48170	ABR48170 Human bla
15	1967.5	48.5	418	6	ABU56635	ABU56635 Lung can
16	1492	36.8	306	3	AA076247	AA076247 Fragment
17	1492	36.8	306	3	AA076247	AA076247 Fragment
18	1039	25.6	229	5	AA09746	AA09746 Amino aci
19	465	11.5	98	4	AA041769	AA041769 Human pol
20	297	7.3	1212	2	AA087503	AA087503 Human N-m
21	297	7.3	1938	6	ABP76679	ABP76679 Streptomy
22	295.5	7.3	1061	2	AA087504	AA087504 Human N-m
23	292	7.2	1081	2	AA087504	AA087504 Human N-m
24	292	7.2	1081	5	AB056223	AB056223 Human NMD
25	292	7.2	1081	7	AA061440	AA061440 Human N-m
26	292	7.2	1081	7	AA039251	AA039251 Human NMD
27	288.5	7.0	1938	6	ABP76679	ABP76679 Streptomy
28	286	6.9	1938	6	ABP76680	ABP76680 Streptomy
29	285.5	7.0	1232	3	AA026239	AA026239 Human N-m
30	285.5	7.0	1232	5	AA047961	AA047961 Human NMD
31	285.5	7.0	1232	6	ABU61439	ABU61439 Human N-m
32	285.5	7.0	1232	7	AA039250	AA039250 Human NMD
33	274.5	6.8	4315	5	ABP43908	ABP43908 MUC5B par
34	273.5	6.6	1938	6	ABP76678	ABP76678 Streptomy
35	273.5	6.6	1938	6	ABP76682	ABP76682 Streptomy
36	270.5	6.7	1938	6	AB098398	AB098398 Streptomy
37	269.5	6.6	1093	2	AA041001	AA041001 Human myo
38	268	6.5	1938	6	AB098398	AB098398 Streptomy
39	267	6.6	50	6	ADA57202	ADA57202 Human sec
40	267	6.6	50	6	ADA41079	ADA41079 Human sec
41	267	6.6	50	7	AA011753	AA011753 Human sec
42	267	6.6	51	3	AA076131	AA076131 Human sec
43	267	6.6	1938	6	ABP76682	ABP76682 Streptomy
44	255.5	6.3	1938	6	ABP76678	ABP76678 Streptomy
45	253.5	6.2	695	4	ABU53152	ABU53152 Human tes

#### ALIGNMENTS

RESULT 1

AAB08864

ID AAB08864 standard; protein; 626 AA.

XX

AC AAB08864;

XX

DT 15-JAN-2001 (first entry)

XX

DE Amino acid sequence of a human secretory protein.

XX

KW Human; secretory protein; HSECP; cancer; gastrointestinal disorder; inflammation; cardiovascular disorder; neurological disorder.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT Peptide 1..27

FT Modified-site 28 /note= "signal sequence"

FT Modified-site 61 /note= "potential phosphorylation site"

FT Modified-site 118 /note= "potential phosphorylation site"

FT Modified-site 144 /note= "potential phosphorylation site"

FT Modified-site 146 /note= "potential phosphorylation site"

FT Modified-site 188 /note= "potential phosphorylation site"

FT Modified-site 252 /note= "potential phosphorylation site"

FT Modified-site 275 /note= "potential phosphorylation site"

FT

FT	Modified-site	336	"potential phosphorylation site"
FT	Modified-site	410	"potential phosphorylation site"
FT	Modified-site	423	"potential phosphorylation site"
FT	Modified-site	467	"potential phosphorylation site"
FT	Modified-site	501	"potential phosphorylation site"
FT	Modified-site	506	"potential phosphorylation site"
FT	Modified-site	524	"potential phosphorylation site"
FT	Modified-site	536	"potential phosphorylation site"
FT	Modified-site	550	"potential phosphorylation site"
FT	Modified-site	593	"potential phosphorylation site"
FT	Modified-site	598	"potential phosphorylation site"
FT	Modified-site	593	"potential phosphorylation site"

AA WO200052151-A2.

XX  
PD  
08-SEP-2000.XX  
PF 03-MAR-2000; 2000WO-US005621.XX  
PR 05-MAR-1999: 99US-0123117P.

XX PA (INCY-) INCYTE PHARM INC.

Tang YT. Lal P. Baughn MR. Yue H. Au-Young J, Lu DAM, Azimzai Y;  
XX  
XX

XX  
DR  
WPI: 2000-579282/54.

DR N-PSDB; AAA75118.

Twenty two human secretory proteins for diagnosing, treating and preventing cancer, inflammation, and gastrointestinal, cardiovascular and neurological disorders.

XX  
PS  
Claim 1: Page 88-89: 107pp: English.

The present sequence represents a human secretory protein, designated HSRCP-1. The specification also describes HSRCP-2 to HSRCP-22. The proteins are useful for diagnosing, treating and preventing cancer, inflammation, and gastrointestinal, cardiovascular and neurological disorders. The proteins may also be used to identify agonists, antagonists, and inhibitors. The polynucleotides may be used for producing the protein recombinantly, and as a source of probes and primers for isolating and identifying related sequences.

Sequence 626 AA;

Alignment Scores:

Assignment Scores:			
Pred. No.:	4.9e-231	Length:	626
Score:	3335.00	Matches:	626
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	82.22%	Indels:	0
DB:	3	Gaps:	0

US-09-914-958B-35 (1-2177) x AAB08864 (1-626)

QV 123 ATGGCTCCGAGCGCTGACCCCGCATGTCCAGGATGTACCGTTCCCTGCTGCTCTGG 182

db

183 TTTCTGCCCATCACTGAGGGTCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242





QY	567	TTGTTCAAGTTC	CCCAATAAACCGGTGGGAAGACATCTCTGACGATGAGGTCAACGTGGCC	626
Db	141	LeuPheLysPheArgAsnAsnArgT	rgLValAspIleLeuSerAspGluValAsnValAla	160
QY	627	CGTGGTGGCCAGCCTCTT	GGCGAGCTCTGCGCTGTGGCTGTGGCAGAAAGGGCTCT	686
Db	161	ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaLacysValAspArgLysGlySer	180	
QY	687	GGAGCGTACTCTATCTACATTCGCAATTTACGCTACGGTAATGTGGGCCCTGATGCCCTC	746	
Db	181	GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu	200	
QY	747	ATTGAATGGACCTTGAGCCAGTGA	CTCTCCCGGGGCATCTGCGCCTCAGAGATGTG	806
Db	201	IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal	220	
QY	807	GCTGCTGAGGCTGGGTGACAAATATACAGGGGCCGAGGCGTCAGCGTGGGCCCATC	866	
Db	221	AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle	240	
QY	867	CTCAGCAGCAGTCCCTCGGATATCTCTCGCACATGAGATGGCCCTAACCTTCCTTTTC	926	
Db	241	LeuSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe	260	
QY	927	CACAAACCGGGGATGGACACCTTTGTGACGCTGGCGCAGTCTGCTGTGGACGACCC	986	
Db	261	HisAsnArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspPro	280	
QY	987	CACAGCATGGCGAGGTGTGCTCCCTGCTGCTGCTTCAACCGTGTGGCAAGTGGACATC	1046	
Db	281	HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle	300	
QY	1047	GTCTATGGCACTGGAATGGCCCCCACCGCCTCTATCTGCAAAATGAGCACCCATGGGAG	1106	
Db	301	ValTyrGlyAsnTyrAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys	320	
QY	1107	GTCCGCTCCGGGACATCGCTCACCCCAAGTTCTCCATGCCCTCCCTGTCGCGACGTC	1166	
Db	321	ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal	340	
QY	1167	ATCACCGCGCATTGACAAATCACAGGAGCTGGAGATCTTCTTCAACAACTATGCTTAC	1226	
Db	341	IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr	360	
QY	1227	CGCAGCTCTCTACCCCAACCGCTCTTCCCGCTCATCCGTAGAGACACGAGACCCCTC	1286	
Db	361	ArgSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu	380	
QY	1287	ATCGAGGAGCTCAATCCCGGACGCTTGGAGCCTGAGCGCGGGCACAGGGGTGTG	1346	
Db	381	IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal	400	
QY	1347	GTGACCGACTTCGACGAGACGGATGTGGACTCTATCTGTCCCATGGAGAGTCCATG	1406	
Db	401	ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet	420	
QY	1407	GCTCAGCCGCTCTCCCTTCCGGGCAATFCAGGGCTTCAACAACTGCTGCGAGTG	1466	
Db	421	AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal	440	
QY	1467	GTGCCACGACCCGCTTGGGCTTGGCGAGTTCAGGGAGCTAAGTCTGCTCTACACCAAG	1526	
Db	441	ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys	460	
QY	1527	AAGAGTGGGGCCCACTCAGGATCATCGACGGGGCTCAGGCTACCTGTGTAGATGGAG	1586	
Db	461	LysSerGlyAlaHisLeuArgIleAlaAspGlySerGlyTyrLeuCysGluMetGlu	480	
QY	1587	CCCGTGCGCACTTTGGCCTGGGGAAGGATGAAGCCAGAGTGTGAGGTGACGTGGCCA	1646	
Db	481	ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro	500	
QY	1647	GATGCGAAGATGTGAGCCGGAACGTGCGCCAGCGGGAGATGAACCTAGTCTGGAGATC	1706	
Db	501	AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle	520	
QY	1707	CTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCTGAGTGTGGCAAGGA	1766	
Db	521	LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly	540	
QY	1767	TTCTCCGACGAGGAATGGCATTGTCATGCACACCAATGAATGAATCATCCAGTTCCTCATTC	1826	
Db	541	PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe	560	
QY	1827	GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGTACAGGTGCGGACC	1886	
Db	561	ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr	580	
QY	1887	AACAGAAGTCCAGTCCGGGCTACGAGCCCAACGAGGATGACACAGCCTGCGTGGGCTGG	1946	
Db	581	AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyTrp	600	
QY	1947	TGGAGCCCTGTGTAAGATAGTGCACACCAAGTTGGGAAGAGCTTGGTCCC	2000	
Db	601	TrpSerProValLeuLysIleValThrProGlnValGlyLysSerLeuGlyPro	618	
RESULT 3				
ID	AAO18264	AAO18264 standard; protein; 653 AA.		
XX	AAO18264;			
DT	26-SEP-2002	(first entry)		
DE	Human chondrocyte specific CEP-68 protein SEQ ID NO:4.			
XX	Human; chondrocyte specific protein; CEP-68; cartilage; marker;			
KW	extracellular matrix protein.			
OS	Homo sapiens.			
Key	Location/Qualifiers			
Peptide	1..19			
FT	/label= signal_peptide			
FT	20..653			
FT	/label= mature CEP-68			
XX	WO200253709-A1.			
XX	11-JUL-2002.			
XX	24-DEC-2001; 2001WO-EP015307.			
XX	05-JAN-2001; 2001DE-0.000305.			
XX	(CYTO-) CYTONET GMBH & CO KG.			
XX	Richter W, Steck E;			
XX	WPI; 2002-528856/56.			
XX	N-PSDB; AAL47917, AAL47918.			
XX	New nucleic acid encoding an extracellular matrix protein, useful, for example, as a marker for detecting or isolating chondrocytes.			
XX	Claim 1; Page 92-94; 107pp; German.			
XX	The present invention relates to the protein and coding sequences of novel human extracellular matrix proteins, more specifically chondrocyte specific CEP-68 proteins. The sequences can be used in the identification of cell differentiation stages in cells which develop from mesenchymal stem cells, particularly chondrocytes and to isolate chondrocytes. The present sequence is the CEP-68 protein			
XX	Sequence 653 AA;			
SQ				

301	ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyIys	320
1107	GTCCGCTTCGGGACATCGCTCACCCAAAGTTCTCCATGCCTCCCTGCCGACCGGT	1166
321	ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal	340
1167	ATCACCGCCGACTTGTGCATATGACAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC	1226
341	IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr	360
1227	CGAGCTCTTCAGCCAAACCCGCTCTCCGCGTGCATCGTATAGAGACGCGAGACCCCTC	1286
361	ArgSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu	380
1287	ATCAGAGAGCTCAATCCCGCGCGAGCCCTTGGAGCTGAGGCGCCGGGCGCACAGGGGGTGTG	1346
381	IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal	400
1347	GTGACCGACTTCGACGGAGACGGGATGCTGGAACCTCATCTTGTCCCATGGAGATCCATG	1406
401	ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet	420
1407	GCTCAGCGCTGTCCGCTTCCTCCGGGGCAATCAGGGCTTCAACACAACTCGCTCGGAGTG	1466
421	AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal	440
1467	GTGCCACGACCCCGGTTTGGGGCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG	1526
441	ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaIysValValLeuTyrThrIys	460
1527	AAGAGTGGGGCCCACTTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAG	1586
461	LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu	480
1587	CCGCTGGCACACTTGTGCTCCCTGGGGAAGATGAACCGACGACGTGTGGAGTCACTGGGCCA	1646
481	ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerValGluValThrTrpPro	500
1647	GATGGCAAGATGTGACCGCGAACGTGGCCGACGGGGAGATGAACCTCAGTCTCGAGATC	1706
501	AspGlyIysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle	520
1707	CTCTACCCCGGGATGAGGACACACTTTCAGACCCAGCCGCCACCTGGAGTGTGGCCAGGA	1766
521	LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly	540
1767	TTCTCCCGACAGGAANAATGGCCATTGCATGGACACCAATGATGCATCCAGTCCCATTC	1826
541	PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe	560
1827	GTGTGCCCTCGAGACAGCCCGTATGTGTCAACACCTATGAAGCTACAGGTGCCGGACC	1886
561	ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr	580
1887	AACAAGAAGTGCAGTCCGGGCTACGACCCCAACGAGGATGCACAGCCTCGCTGGGCTGG	1946
581	AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly---	599
1947	TGGAGCCCTGTGTTGAAGATAGTAGACACACAGTTCGGGAAGACCTTGTGCCC	2000
600	-----ThrLeuGlyGlnSerProGlyPro	607
RESULT 4		
AAE18679		
ID	AAE18679 standard; protein; 653 AA.	
XX		
XX	AAE18679;	
XX		
XX		
DT	17-MAY-2002 (first entry)	
XX		
DE	Human SCIM-1 protein.	
XX		
XX	Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;	





Db 81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100  
QY 447 GACCGGAGGGAGACCGCATTTGGGTACAGCTCGCATCGAGCGGGAGCGCGGGAG 506  
Db 101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120  
QY 507 GAGATCTACTTCTCAACACCAATAATGCTTCTCGGGGGTGGCCACGCTACACCGACAAAG 566  
Db 121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140  
QY 567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAACGTGGCC 626  
Db 141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160  
QY 627 CGTGTGTGGCCAGCCTCTTTGCCGAGCGCTCTGTGGCTGTGTGTGGACAGAAAGGCTCT 686  
Db 161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180  
QY 687 GGAGCTACTCTATCTACATTGCCAATTACGGCTACGGTAAATGTGGCCCTGATCCCTC 746  
Db 181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200  
QY 747 ATTGAATGGACCTGAGCGCCAGTCACCTCTCCCGGGCATTTCTGGCGCTCAGAGATGG 806  
Db 201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220  
QY 807 GCTGTGAGGCTGGGTGAGCAATAATACAGGGGGCCGAGCGCTCAGGTGGGCCCATC 866  
Db 221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240  
QY 867 CTCAGCAGAGTGCCTCGGATATCTCTCGCAATAGAAATGGCGCTAACTTCCTTTTC 926  
Db 241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260  
QY 927 CACACCGGGGGGATGGACACCTTTGTGACGCTGCGCGCAGTGTGGTGTGGACACCC 986  
Db 261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280  
QY 987 CACAGCATGGCGAGGTGTGCGCTGTGCTGCTCAACCTGATGGCAAGTGCACATC 1046  
Db 281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300  
QY 1047 GTCTATGGCACTGGAATGGCCCCCACCGCCTCTATCTGCAATGAGCACCCATGGGAAG 1106  
Db 301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320  
QY 1107 GTCCGCTTCCGGGACATCGCCTCACCCAAATTTCTCATGCCCTCCCTGTTCGCGACGTC 1166  
Db 321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340  
QY 1167 ATCAGCGCGACTTTGACAAATCACAGGAGCTGGAGATCTTCTCAACAACTTGCTAC 1226  
Db 341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360  
QY 1227 CGCAGCTCTCTAGCCAAACCGCTCTTCCCGCTCATCCGTPAGAGACGCGAGACCCCTC 1286  
Db 361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380  
QY 1287 ATCAGGAGCTCAATCCCGGCGACGCTTGGAGCTGAGGCGCGGGGACAGGGGTGTG 1346  
Db 381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400  
QY 1347 GTGACCGACTTCGAGGACGCGGATGCTGGACCTCATCTTGTCCCATGAGAGTCCATG 1406  
Db 401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420  
QY 1407 GCTAGCGCGCTCTCGCTCTTCCGGGCAATCAGGGCTTCAACAACTTGGCTGCGAGTG 1466  
Db 421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440  
QY 1467 GTGCCACGACCCGGTTTGGGGCTTTGCCAGGGAGCTAAAGTGTGTCTTACACCAAG 1526  
Db 441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

QY 1527 AAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586  
Db 461 LysSerGlyAlaHisLeuArgIleIleAspGlyLysGlySerGlyTyrLeuCysGluMetGlu 480  
QY 1587 CCGTGGCACACTTTTGGCTGGGAAGATGAAGCAGCAGTGTGGAGGTGACGTGGCCA 1646  
Db 481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500  
QY 1647 GATGCAAGATGTGAGCGGGAACGTGCCACGCGGGAGATCAACTCAGTGTGGAGATC 1706  
Db 501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520  
QY 1707 CTCACCCCGGATGAGGACACACTTTCAGGACCCAGCCCACTCGAGTGTGCCAAGGA 1766  
Db 521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540  
QY 1767 TTCTCCAGCAGGAAATGGCATTGTGATGACACCAATGAATGATCCAGTTCCTCATTC 1826  
Db 541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560  
QY 1827 GTGTCCCTCGAGACAGCCCGTATGTGTCAACCTATGGAGCTACAGGTGCCGACC 1886  
Db 561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580  
QY 1887 AACAGAAAGTGCAGTCGGGGCTTACGAGCCCAACAGGATGGCACAGCCTCGCTGGGCTGG 1946  
Db 581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599  
QY 1947 TGGAGCCCTGTGTTGAAGATAGTACACCAAGTTGGGAAGAGCCTTGGTCCC 2000  
Db 600 -----ThrLeuGlyGlnSerProGlyPro 607  
RESULT 6  
ABU56636  
ID ABU56636 standard; protein; 653 AA.  
XX  
AC ABU56636;  
XX  
DT 02-APR-2003 (first entry)  
XX  
DE Lung cancer-associated polypeptide #299.  
XX  
KW Lung cancer-associated polypeptide; cytostatic; emphysema;  
KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
XX  
OS Unidentified.  
XX  
PN WO200286443-A2.  
XX  
PD 31-OCT-2002.  
XX  
PF 18-APR-2002; 2002WO-US012476.  
XX  
PR 18-APR-2001; 2001US-0284770P.  
PR 10-MAY-2001; 2001US-0290492P.  
PR 09-NOV-2001; 2001US-0339245P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 29-NOV-2001; 2001US-034370P.  
PR 12-APR-2002; 2002US-0372246P.  
XX  
PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
PI Aziz N, Murray R;  
XX  
DR WPI; 2003-093161/08.  
DR N-PSDB; ABX76365.  
XX  
PT Detecting a lung cancer-associated transcript in a cell from a patient  
for treating a lung cancer, by contacting a biological sample from the





Db	541	PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe	560
Qy	1827	GTGTGGCCCTCGAGACAGCCGGTATGTCACACACCTATGAAAGCTACAGTGCCGGACC	1886
Db	561	ValCysProArgaspIysProValCysValAsnThrTyrGlySerTyrA-9CysA-9gThr	580
Qy	1887	AACAAGAAAGTCAGTCGGGGCTACAGACCCCAACAGAGATGGCACAGCTCGTGGGCTGG	1946
Db	581	AsnIysLysCysSerA-9GlyTyrGluProAsnGluAspGlyThrAlaCysValGly---	599
Qy	1947	TGGAGCCCTGTGTTGAAGATGATGACACCAAGCTGGGAAGACCTTGTCTCC	2000
Db	600	-----ThrLeuGlyGlnSerProGlyPro	607
RESULT 7			
AAE18680			
ID	AAE18680	standard; protein; 637 AA.	
XX	AC	AAE18680;	
XX	AC		
XX	DT	17-MAY-2002 (first entry)	
XX	XX	Human SCIM-1 splice variant A protein.	
XX	XX	Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;	
KW	KW	extracellular matrix protein; tolerogenic protein; Grave's disease;	
KW	KW	autoimmune disorder; juvenile arthritis; primary glomerulonephritis;	
KW	KW	polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;	
KW	KW	rheumatoid arthritis; Addison's disease; primary biliary sclerosis;	
KW	KW	uveitis; systemic lupus erythematosus; inflammatory bowel disease;	
KW	KW	multiple sclerosis; diabetes; inflammatory disorder; therapy.	
OS	OS	Homo sapiens.	
XX	XX	WO200206478-A2.	
FN	FN	24-JAN-2002.	
PD	PD		
XX	XX	09-JUL-2001; 2001WO-EP007888.	
PF	PF		
XX	XX	13-JUL-2000; 2000EP-00202495.	
PR	PR	(ALKU ) AKZO NOVEL NV.	
XX	XX		
PA	PA		
XX	XX	Heus HC, Nelissen RLH, Meeuwisse CML;	
EI	EI		
XX	XX	WPI; 2002-179796/23.	
DR	DR	N-PSDB; AAD29791.	
XX	XX		
PT	PT	New extracellular matrix protein useful for prevention of inflammatory	
PT	PT	diseases, more specifically in induction of T cell tolerance to the	
XX	XX	protein in patients suffering from rheumatoid arthritis.	
PS	PS	Claim 6; Page 38-40; 45pp; English.	
XX	XX		
CC	CC	The invention relates to human extracellular matrix protein i.e	
CC	CC	Synovium/cartilage inflammation-linked Messenger-1 (SCIM-1) and nucleic	
CC	CC	acid molecules encoding such proteins. Sequences of the invention are	
CC	CC	useful in a screening assay for the identification of tolerogenic	
CC	CC	polypeptides. SCIM proteins are useful in therapy and for manufacture of	
CC	CC	a pharmaceutical preparation against inflammatory diseases and for	
CC	CC	induction of immunological tolerance to an autoantigen in patients	
CC	CC	suffering from autoimmune disorders, more specifically rheumatoid	
CC	CC	arthritis. The inflammatory disorders treated include diseases like	
CC	CC	Grave's diseases, juvenile arthritis, primary glomerulonephritis,	
CC	CC	polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,	
CC	CC	rheumatoid arthritis, Addison's disease, primary biliary sclerosis,	
CC	CC	uveitis, systemic lupus erythematosus, inflammatory bowel disease,	
CC	CC	multiple sclerosis and diabetes. Polypeptides of the invention are also	
CC	CC	useful in a diagnostic method for the detection of activated autoreactive	
XX	XX	T cells. The present sequence is human SCIM-1 splice variant A protein	



Db ||||| PheGluIleValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80  
QY 387 CAGAAGCGCTGGTGAACATCGCGGTGATGAGCGAGCTACCTACCTACGCGCTGCGG 446  
Db 81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100  
QY 447 GACCGCAGGGAAACCCATTGGGTCCACAGCTGCGAGCATCGACGGGACGCCGGGAG 506  
Db 101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120  
QY 507 GAGATCTACTTCTCAACACCATAATGCTTCTCGGGGTGGCCACGTACACCCGACAAG 566  
Db 121 GluIleTyrPheLeuAsnThrAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140  
QY 567 TTGTTCAGTTCCGAATTAACCGGTGGGAAGACATCTGAGCGCATGAGGTCAACGTGGCC 626  
Db 141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160  
QY 627 CGTGGTGTGCCAGCCTCTTTGCCGACGCTCTGTGGCTGTGTGACAGAAAGGCTCT 686  
Db 161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180  
QY 687 GAGCGTACTCTATCTACATTGCCAATTACGCTACGGTAAATGTGGGCCCTGATGCCCTC 746  
Db 181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200  
QY 747 ATGGAATGACCTGAGCGCAGTGACCTTCCGGGGCATCTTCCGGCTCAGAGATGTG 806  
Db 201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220  
QY 807 GCTGCTGAGCTGGGTGACGAAATATACAGGGGGCCGAGCGTCAAGCTGGGCCCCATC 866  
Db 221 AlaAlaGluAlaGlyValSerLysTyrThrGlyArgGlyValSerValGlyProIle 240  
QY 867 CTCAGCAGAGTGCCTCGGATATCTTCTCGCAATCAGAAATGAGTGGCTAACTTCTCTTC 926  
Db 241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260  
QY 927 CACAACCGGGGATGGACCTTTGTGGAGCTGTGGCGCCAGTCTGTGTGGACGACCCC 986  
Db 261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspPro 280  
QY 987 CACACGATGGCGAGCTGTGCGCTGCTGACTTCAACCGTGTGCAAGTGGACATC 1046  
Db 281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300  
QY 1047 GTCTATGGCACTGGGAATGCCCCACCCGCTCTATCTGCAATGAGCACCCATGGGAAG 1106  
Db 301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320  
QY 1107 GTCCGCTTCGGGACATCGCCTCACCAAGTTCTCCATGCTCCCTCCCTGTCGCGAGTTC 1166  
Db 321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340  
QY 1167 ATCACC CGCGCATTTGACAAATCACAGGAGCTGGAGATCTTCTCAACAACATGGCTAC 1226  
Db 341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360  
QY 1227 CGCAGCTCTCTAGCCAAACCGCTCTTCCGCGTATCCGTTAGAGAGCACGAGACCCCTTC 1286  
Db 361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380  
QY 1287 ATCGAGGAGCTAATCCCGCGCAGCGCTTGGAGCTTGAGCGCGGGGCACAGGGGTGG 1346  
Db 381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400  
QY 1347 GTGACCGACTTCGACGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGATCCATG 1406  
Db 401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420  
QY 1407 GCTCAGCGCTGTCCCTCTTCCGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTGT 1466

421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440  
QY 1467 GTGCCACGACACCGGTTTGGGGCTTTGGCAGGGGAGCTTAAGTCTGCTCTACACCAAG 1526  
Db 441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460  
QY 1527 AAGAGTGGGGCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586  
Db 461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480  
QY 1587 CCGGTGGCACACTTTGGCTGGGGAAGGATGAAGCAGCAGTGTGGAGGTGACGTGGCCA 1646  
Db 481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500  
QY 1647 GATGGCAAGATGTTGAGCCGGAACCTGGCCAGCGGGGAGATGACTCAGTCTGGAGATC 1706  
Db 501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520  
QY 1707 CTCTACCCCGGATGAGGACACACTTACGAGCCACGAGCCCTGAGTGTGGCCCAAGGA 1766  
Db 521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaPheLeuGluCysGlyGlnGly 540  
QY 1767 TTCTCCACGACAGAAATGGCCATTCATGATGACACCAATGAATGCAATGCCATTC 1826  
Db 541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560  
QY 1827 GTGTGCTCGAGACAGAGCCGTTATGTGTCAACACTATGGAAGCTACAGGTGCGGACC 1886  
Db 561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580  
QY 1887 AACCAAGATGCACTCGGGCTACGAGCCCAACGAGAGTGGCACAGCTGCGGTG 1940  
Db 581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysVal 598

RESULT 9  
AAO18267  
ID AAO18267 standard; protein; 636 AA.  
XX AAO18267;  
XX  
XX 26-SEP-2002 (first entry)  
XX Human chondrocyte specific CEP-68 protein SEQ ID NO:25.  
DE Human; chondrocyte specific protein; CEP-68; cartilage; marker;  
XX extracellular matrix protein.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 377 /note= "encoded by GGA"  
FT Misc-difference 433 /note= "encoded by TTC"  
FT Misc-difference 541 /note= "encoded by TTC"  
FT Misc-difference 550 /note= "encoded by ATG"  
FT Misc-difference 557 /note= "encoded by CAGTTC"  
XX  
XX WO200253709-A1.  
XX  
XX 11-JUL-2002.  
XX  
XX 24-DEC-2001; 2001WO-EP015307.  
XX  
XX 05-JAN-2001; 2001DE-01000305.  
XX (CYTO-) CYTONET GMBH & CO KG.  
XX  
XX Richter W, Steck E;  
XX



Db 580 AsnlylsysCysSerArgGlyTyrGluProAnGluAspGlyThrAlaCysValAlaGln 1976  
 Qy 1941 -----GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976  
 |||::: ||| ::::| |||  
 Db 600 ValAlaPheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 616  
 RESULT 10  
 ABR48172  
 XX ABR48172 standard; protein; 1597 AA.  
 AC ABR48172;  
 XX  
 XX 12-JUN-2003 (first entry)  
 DT  
 XX Human bladder cancer associated protein sequence SEQ ID NO:60.  
 XX Human; bladder cancer; cytostatic; gene therapy; vaccine.  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO2003003906-A2.  
 FN  
 XX  
 XX 16-JAN-2003.  
 PD  
 XX  
 XX 03-JUL-2002; 2002WO-US021338.  
 PF  
 XX  
 XX 03-JUL-2001; 2001US-0302814P.  
 PR  
 XX 03-AUG-2001; 2001US-0310099P.  
 PR 08-NOV-2001; 2001US-0343705P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 12-APR-2002; 2002US-0372246P.  
 PR  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 PA  
 XX  
 XX Mack DH, Aziz N;  
 FI  
 XX  
 XX WPI; 2003-201532/19.  
 DR  
 XX N-PSDB; ACC50983.  
 DR  
 XX  
 XX Detecting a bladder cancer-associated transcript in a cell from a  
 PT patient, comprises contacting a biological sample from the patient with a  
 PT bladder cancer-associated polynucleotide or antibody.  
 XX  
 XX Claim 10; Page 250; 307pp; English.  
 PS  
 XX  
 CC The present invention describes a method for detecting a bladder cancer-  
 CC associated transcript in a cell from a patient. The method comprises  
 CC contacting a biological sample from the patient with a polynucleotide  
 CC that selectively hybridises to a sequence that is 80 % identical to a  
 CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059  
 CC encode the human bladder cancer-associated proteins given in ABR48146 to  
 CC ABR48242). Bladder cancer-associated sequences from the present invention and  
 CC have cytostatic activities, and can be used in antisense gene therapy and  
 CC in vaccine production. The method can be used for detecting a bladder  
 CC cancer-associated transcript in a cell from a patient. The method is  
 CC useful in diagnosing or treating bladder cancer and in screening for  
 CC compounds that modulate bladder cancer, such as hormones or antibodies.  
 CC The nucleic acid molecules from the present invention may be used in  
 CC various screening and diagnostic methods, and for gene therapy, vaccine  
 CC and/or antisense/inhibition applications  
 CC  
 XX Sequence 1597 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3 95e-159 Length: 1597  
 Score: 2337.00 Matches: 566  
 Percent Similarity: 35.18% Conservative: 6  
 Best Local Similarity: 34.81% Mismatches: 29  
 Query Match: 57.62% Indels: 1025  
 DB: 6 Gaps: 11

Db 327 spGlyGlnPheLysGluGluAlaAlaLeuValGluGluArgGluAlaGlyAla 347  
 QY 836 ----- 836  
 Db 347 laGlyValProArgGlyArgValArgThrAlaLeuGlnThrSerLysSerHisLeuAla 367  
 QY 836 ----- 836  
 Db 367 spLysAsnLeuPheGlyProProCysTyrTyrSerValCysAlaProSerProAlaHisP 387  
 QY 836 ----- 836  
 Db 387 roPheProAlaArgGlnAlaProGlnHisTyrProValAlaProLeuValThrGlnLeuM 407  
 QY 836 ----- 836  
 Db 407 etThrHisGlyArgLeuAlaGlyLysLeuAlaArgSerValProHisProArgAlaProG 427  
 QY 836 ----- 836  
 Db 427 lyMetAspProLysCysLysGlyArgHisAlaGluProGlyLeuMetAlaGluAlaLeuG 447  
 QY 836 ----- 836  
 Db 447 lyAlaTrpProAlaLeuSerThrThrValValProGlyGlyLeuArgSerTrpGluGluS 467  
 QY 836 ----- 836  
 Db 467 erArgGlnLysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTrpS 487  
 QY 836 ----- 836  
 Db 487 erClnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProProIleL 507  
 QY 836 ----- 836  
 Db 507 euGlnArgThrAspGlyAspProGlyArgArgAspSerProLysValThrGlnGluC 527  
 QY 836 ----- 836  
 Db 527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547  
 QY 836 ----- 836  
 Db 547 ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP 567  
 QY 836 ----- 836  
 Db 567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587  
 QY 836 ----- 836  
 Db 587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607  
 QY 836 ----- 836  
 Db 607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAla 627  
 QY 836 ----- 836  
 Db 627 rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647  
 QY 836 ----- 836  
 Db 647 euArgGlySerProValLeuGlnValGlyLeuAlaSerAlaThrHisCysGlyS 667  
 QY 837 -----GGGGCCGAGCGTCAGCGTGGGCCCCATCTCAGCAGCATGCGCT 882  
 Db 667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerAlaAs 687  
 QY 883 CGGATATCTTCGACAAATAGAAATGGGCTTAATCTTCCTTCCACACCGGGCGCATG 942  
 |||||

Db 687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707  
 QY 943 GCACCTTTGTGAGCGTGGCCAGTGCT----- 971  
 Db 707 lyThrPheValAspAlaAlaSerAlaGluArgArgLeuAlaPheIleValHisLeuL 727  
 QY 971 ----- 971  
 Db 727 ySTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS 747  
 QY 971 ----- 971  
 Db 747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisGlyL 767  
 QY 971 ----- 971  
 Db 767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787  
 QY 971 ----- 971  
 Db 787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807  
 QY 971 ----- 971  
 Db 807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrIleValL 827  
 QY 971 ----- 971  
 Db 827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847  
 QY 972 -----GGTGTGACGACCCACCAGCATGGCGAGGTGTGCGCTGGCTGACTTCAACC 1026  
 Db 847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867  
 QY 1027 GTGATGCGCAAGTGGACATCGTCTATGGCACTGGAAATGGCCCCCGCTCTATCTGC 1086  
 Db 867 rgaspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887  
 QY 1087 AATAGACACCCATGGAAAGTCCGCTTCGGGACATCGCTCACCCAGTTCCTCATGC 1146  
 Db 887 inMetSerThrHisGlyLysValArgPheArgaspIleAlaSerProLysPheSerMetP 907  
 QY 1147 CCTCCCTTCGCCACGTCATCACCGCGACTTTGACAAATGACAGGACTGGAGATCT 1206  
 Db 907 roSerProValArgThrValIleThrAlaAspPheAspAsnaspGlnGluLeuGluIleP 927  
 QY 1207 TCTTCAACACATTCCTACCGAGCTCTCAGCCAAACCGCTCTTCCGC----- 1256  
 Db 927 hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL 947  
 QY 1256 ----- 1256  
 Db 947 euAlaArgGlySerSerSerLeuThrAlaGlyArgAsnGlyGlnGlyGluGlyLeuA 967  
 QY 1256 ----- 1256  
 Db 967 rgIleArgArgGlyGlyPheProGlyProGlyGlyClnAlaLysValAsnThrGlyProL 987  
 QY 1256 ----- 1256  
 Db 987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007  
 QY 1256 ----- 1256  
 Db 1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA 1027  
 QY 1256 ----- 1256  
 Db 1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047  
 QY 1256 ----- 1256  
 Db 1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067

QY 1256 ----- 1256  
 Db 1067 ySGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrA 1087  
 QY 1256 ----- 1256  
 Db 1087 rGLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107  
 QY 1256 ----- 1256  
 Db 1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlns 1127  
 QY 1256 ----- 1256  
 Db 1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147  
 QY 1256 ----- 1256  
 Db 1147 rolleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167  
 QY 1256 ----- 1256  
 Db 1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187  
 QY 1256 ----- 1256  
 Db 1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207  
 QY 1256 ----- 1256  
 Db 1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227  
 QY 1256 ----- 1256  
 Db 1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247  
 QY 1256 ----- 1256  
 Db 1247 rGLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267  
 QY 1256 ----- 1256  
 Db 1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlns 1287  
 QY 1257 -----GTCATCCGTAGAGACCGAGACCGAGACCC 1284  
 Db 1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307  
 QY 1285 TCATCGAGGAGCTCAATCCGGCGAGCCCTCGAGGCTCGAGGCGCGGGGCACAGGGGGTG 1344  
 Db 1307 euileGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327  
 QY 1345 TGGTGACCGACTTCGACGAGACGGGATGCTGACCTCATCTGTCCCATGGAGAGTCCA 1404  
 Db 1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347  
 QY 1405 TGGCTCAGCGCTGTCCTGCTTCCTCGGGGCATCAGGGCTTCAACACACTGGCTCGGAG 1464  
 Db 1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnTrpLeuArgV 1367  
 QY 1465 TGGTGCCAGCCACCGGTTGGGGCTTTGCCAGGGGAGCTAAGGTGCTCTACACCA 1524  
 Db 1367 alValProArgThrArgPheGlyAlaPheAlaargGlyAlaLysValLeuTyrThrL 1387  
 QY 1525 AGAAGAGTGGGGCCACCTCAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584  
 Db 1387 ySLysSerGlyAlaHisLeuArgIleIleaspGlyGlySerGlyTyrLeuCysGluMetG 1407  
 QY 1585 AGCCCGTGGCACACTTTGGCTCGGGGAGGATGAAGCCAGCTGTGGAGGTGACGTGGC 1644  
 Db 1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrp 1427

QY 1645 CAGATGGCAAGATGTTGAGCCGGGAACCTGGCCACCGGGGAGATGAATCAGTGTGGAGA 1704  
 Db 1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447  
 QY 1705 TCCTCTACCCCGGGATGAGACACACTTCAGGACCCAGCCCACTGGAGTGTGCCAAG 1764  
 Db 1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467  
 QY 1765 GATTCCTCCAGCAGCAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824  
 Db 1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProP 1487  
 QY 1825 TCGTGTGCTTCGAGACAAAGCCGCTATGTGTCAACACCTATGGAAGCTACAGTCCCGGA 1884  
 Db 1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507  
 QY 1895 CCAACAGAAGTGCAGTCGGGGCTACGAGCCACGAGGATGGACACGCTGCGTGGGG- 1943  
 Db 1507 hrAsnLysCysSerArgGlyTyrGluProAsnGluaspGlyThrAlaCysValGlyT 1527  
 QY 1944 -----TGGTGGAGCCCTGTGTGAAG----- 1964  
 Db 1527 hrGluLeuGlySerArgHisThrMetThrTyrLysProArgProLysLysGluLeuGlnL 1547  
 QY 1965 -----ATAGTGACACACAAAG----- 1980  
 Db 1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566  
 QY 1981 TTGGGAAGAGCTTGGTCCCTGAATCACTGCTTGAATCACCCTGGAATA 2040  
 Db 1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586  
 QY 2041 CCTGTGTATCAG 2052  
 Db 1587 ProGluAlaGln 1590  
 RESULT 11  
 ABU56637  
 ID ABU56637 standard; protein; 1597 AA.  
 XX AC ABU56637;  
 XX DT 02-APR-2003 (first entry)  
 XX DE Lung cancer-associated polypeptide #230.  
 XX KW Lung cancer-associated polypeptide; cytostatic; emphysema;  
 KW antinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 XX OS Unidentified.  
 XX WO200286443-A2.  
 XX PN 31-OCT-2002.  
 XX PD 18-APR-2002; 2002WO-US012476.  
 XX PF 18-APR-2001; 2001US-0284770P.  
 XX PR 10-MAY-2001; 2001US-0290492P.  
 PR 09-NOV-2001; 2001US-0339245P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-0334370P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX (BOSB-) EOS BIOTECHNOLOGY INC.  
 XX PA Aziz N, Murray R;  
 XX WI; 2003-093161/08.  
 XX N-PSDB; ABX76366.  
 DR





Db 487 erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProIleL 507  
 QY 836 ----- 836  
 Db 507 euGlnArgThrAspGlyAspProGlyArgArgAspSerProIysValThrGlnGluC 527  
 QY 836 ----- 836  
 Db 527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547  
 QY 836 ----- 836  
 Db 547 ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP 567  
 QY 836 ----- 836  
 Db 567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587  
 QY 836 ----- 836  
 Db 587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607  
 QY 836 ----- 836  
 Db 607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA 627  
 QY 836 ----- 836  
 Db 627 rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647  
 QY 836 ----- 836  
 Db 647 euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlyS 667  
 QY 837 -----GGGGCCGAGCGTCAGCGTGGGGCCCATCTCCAGCAGCATGCGCT 882  
 Db 667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerAlaAs 687  
 QY 883 CGGATATCTTCGCACATGAGATGGCTTAAGTCTCTTCCACACCGGGCGGATG 942  
 Db 687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707  
 QY 943 GCACCTTTGTGAGCGTCGGCGCAGTGCT----- 971  
 Db 707 lyThrPheValAspAlaAlaAlaSerAlaGluArgArgLeuAlaPheIleValHisLeuL 727  
 QY 971 ----- 971  
 Db 727 ysTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyPros 747  
 QY 971 ----- 971  
 Db 747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisGlyL 767  
 QY 971 ----- 971  
 Db 767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787  
 QY 971 ----- 971  
 Db 787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807  
 QY 971 ----- 971  
 Db 807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL 827  
 QY 971 ----- 971  
 Db 827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847  
 QY 972 -----GGTGTGACGACCCCGACCATGGCGAGGTCTCGCCCTGGCTGACTTCAACC 1026  
 Db 847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867

QY 1027 GTGATGGCAAGTGCACATCGTCTATGCAACTGGAATGGCCCGCCACCGCTCTATCTGC 1086  
 Db 867 rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887  
 QY 1087 AAATGAGCACCACCATGGGAAGTCCGCTTCGGGACATCGCCTCACCCAGTTCTCCATGC 1146  
 Db 887 lnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProIysPheSerMetP 907  
 QY 1147 CCTCCCTGTCGGCAGCGTCATCACCCCGAGCTTTGACAATGACACGAGCTGGAGATCT 1206  
 Db 907 roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP 927  
 QY 1207 TCTTCAACAACATGCTTACCGCAGCTCCTCAGCAACCGCTCTCTCCGC----- 1256  
 Db 927 hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL 947  
 QY 1256 ----- 1256  
 Db 947 euAlaArgGlySerSerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA 967  
 QY 1256 ----- 1256  
 Db 967 rgIleArgArgGlyGlyPheProGlyProGlyGlyGlnAlaLysValAsnThrGlyProL 987  
 QY 1256 ----- 1256  
 Db 987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007  
 QY 1256 ----- 1256  
 Db 1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA 1027  
 QY 1256 ----- 1256  
 Db 1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047  
 QY 1256 ----- 1256  
 Db 1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067  
 QY 1256 ----- 1256  
 Db 1067 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1087  
 QY 1256 ----- 1256  
 Db 1087 rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107  
 QY 1256 ----- 1256  
 Db 1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS 1127  
 QY 1256 ----- 1256  
 Db 1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147  
 QY 1256 ----- 1256  
 Db 1147 roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167  
 QY 1256 ----- 1256  
 Db 1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187  
 QY 1256 ----- 1256  
 Db 1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207  
 QY 1256 ----- 1256  
 Db 1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227

```
QY 1256 ----- 1256
Db 1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247
QY 1256 ----- 1256
Db 1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267
QY 1256 ----- 1256
Db 1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlns 1287
QY 1257 -----GTCATCCCTAGACAGACGAGACCCCC 1284
Db 1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307
QY 1285 TCATCGAGGAGCTCAATCCCGGGACACCCCTTGAGCCCTGAGGGCCGGGCACAGGGGGTG 1344
Db 1307 euIleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327
QY 1345 TGGTGACCGACTCGACGGAGAGCGGAGTGTGGACCTCATCTTGTCCATGGAGAGTCCA 1404
Db 1327 alValThrAspPheAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347
QY 1405 TGGCTCAGCGCTGTCGCTTCGCGGGCAATCAGGGCTTCAACAACTGCTGCGAG 1464
Db 1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgV 1367
QY 1465 TGGTGCCACGACCCCGCTTTGGGGCCCTTTCGCCAGGGGAGCTAAGTGTGCTCTTACACCA 1524
Db 1367 alValProArgThrArgPheGlyAlaPhealargGlyAlaLysValLeuTyrThrL 1387
QY 1525 AGAAGAGTGGGGCCCACTGAGGATCATCAACGGGGCTCAGGCTACCTGTGTGAGATGG 1584
Db 1387 ysLysSerGlyAlaHisLeuArgIleLeuAspGlyGlySerGlyTyrLeuCysGluMetG 1407
QY 1585 AGCCCTGGGCACACTTGGCTCGGGAAGATCAAGCCAGCTGTGAGGTGACGTGGC 1644
Db 1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpP 1427
QY 1645 CAGATGCCAAGATGTGTAGCGGAAGCTGGCCAGCGGGAGATGAATCACTAGTCTCTGAGA 1704
Db 1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447
QY 1705 TCTCTACCCCGGATGAGGACACACTTCAGACCCAGCCCACTGGAGTGTGGCAAG 1764
Db 1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467
QY 1765 GATTTCTCCAGCAGGAATGGCCATTGGCATGGACACCAATGAATGCATCCAGTTCCTCAT 1824
Db 1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProP 1487
QY 1825 TGTGTGCTCCCGAGACAAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCGGGA 1884
Db 1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507
QY 1885 CCAACAGAGAGTCATCGGGCTACAGCCCAACAGAGATGCCACAGCTGGTGGGC- 1943
Db 1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyT 1527
QY 1944 -----TGTGGAGCCCTGTGTGAAG----- 1964
Db 1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547
QY 1965 -----ATAGTGACACCAACAAG----- 1980
Db 1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566
QY 1981 TTGGGAAGACCTTGGTCCCTGATCACTGAATCACTGCTGAATCACTGGCTGAATA 2040
Db 1567 LeuLysArgAlaGlnLeuGlnAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586
QY 2041 CCTGTTGATCAG 2052
```

```
Db 1587 ProGluAlaGln 1590
RESULT 12
AAM39983
ID AAM39983 standard; protein; 435 AA.
XX
AC AAM39983;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3128.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-0048725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 13-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
(PHYSE-) HYSRC INC.
XX
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
Zhou P, Goodrich R, Drmanac RT;
WFI; 2001-442253/47.
DR N-PSDB; AAI59139.
XX
Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
XX
Example 4; SEQ ID NO 3128; 10078pp; English.
XX
The invention relates to human nucleic acids (AAI57798-AAI61369) and the
encoded polypeptides (AAM38642-AAM42213) with nootropic,
immunosuppressant and cystostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
Activ/inhibit activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening
assays for receptor activity, arthritis and inflammation, leukaemia and
C.N.S disorders. Note: The sequence data for this patent did not form
part of the printed specification
XX
SQ Sequence 435 AA;
Alignment Scores:
Pred. NO.: 4,69e-144 Length: 435
Score: 2124.50 Matches: 400
```

Percent Similarity: 96.87% Conservatives: 2  
Best Local Similarity: 96.39% Mismatches: 6  
Query Match: 52.38% Indels: 7  
DB: 4 Gaps: 1

US-09-914-958b-35 (1-2177) x AAM39983 (1-435)

QY 753 ATGACCTTGAGCGCAGTACCTCTCCCGGGCATCTTGGCGCTCAGAGATGTGCTGCT 812  
Db 1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20  
QY 813 GAGCGTGGGTCAGCAATATACAGCGGCGCGAGCGTCAGCGTGGGCGCCCATCTCAGC 872  
Db 21 GluAlaGlyValSerLysIleThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40  
QY 873 AGCAGTGCCTGGATATCTTCGACCAATGAGAAATGGCCCTAACTTCTTTTCCACAAC 932  
Db 41 SerSerAlaSerAspIlePheCysAspAsnGluAenGlyProAsnPheLeuPheHisAsn 60  
QY 933 CGGGCGATGGACCTTTGTGACCGCTGGCGCAGTGTGCTGTGGAGCAGCCGCCACGAG 992  
Db 61 ArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspProHisGln 80  
QY 993 CATGGCGAGGTGTCGCCCTGGCTCACTTCAACCGTGCATGGCAAGTGGACATCGCTCTAT 1052  
Db 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyValAspIleValTyr 100  
QY 1053 GCACACTGGAATGGCCCCACCGCTCTATCTGCAAAATAGACACCCATGGGAAGTCCGC 1112  
Db 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120  
QY 1113 TTCGGGACATCGCTCACCAGTCTCCATGCCCTCCCTGTCGGCAGCGTATCACC 1172  
Db 121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140  
QY 1173 GCGCACTTTGACAAAGCAGGAGTGGAGATCTTCTCAACAAATTCCTACCGCCAGC 1232  
Db 141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160  
QY 1233 TCCTCAGCAACCGCTCTTCGCGTATCCGTAGAGACACGAGACCCCTCATCGAG 1292  
Db 161 SerSerAlaAsnArgLeuPheArgValIleArgGluHisGlyAspProLeuIleGlu 180  
QY 1293 GAGCTCAATCCGGCGAGCGCTTGGAGCTGAGGCGCGGCGACAGGGGTGTGTGACC 1352  
Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyValValThr 200  
QY 1353 GACTTCGACGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412  
Db 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220  
QY 1413 CCGCTGTGCTTCCCGGGCAATCAGGCTTCAACAACTGCTGGAGTGGTGCCCA 1472  
Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnTrpLeuArgValValPro 240  
QY 1473 CCACCCGGTTTGGGCTTCCCGGGAGCTAGGTGCTGCTACACCAAGAGAGT 1532  
Db 241 ArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260  
QY 1533 GGGGCGCACCTGAGATCATCGCGGGGCTCAGGCTACTGTGTGAGATGGAGCCGCTG 1592  
Db 261 GlyAlaHisLeuArgIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280  
QY 1593 GCACATTTGGCTGGGGAAGATGAAGCAGCAGTGTGAGGTGAGCTGGCGCAGATGCG 1652  
Db 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300  
QY 1653 AGATGCTGAGCGGAACTGGCGAGCGGGAGATCAACTCAGTGTGAGATCCCTCTAC 1712  
Db 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320  
QY 1713 CCCCGGATGAGGACACACTTACGACCCGAGCCCACTGGAGTGTGCCAAGATTCTCC 1772

Db 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGlyPheSer 340  
QY 1773 CAGCAGGAATGCGCATTCGATCGACACCAATGAATGCATCCAGTTCCTCGTGTGC 1832  
Db 341 GlnGlnGluAenGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCys 360  
QY 1833 CCTCAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGTGCCTGGACCAACAAG 1892  
Db 361 ProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLys 380  
QY 1893 AAGTGCATCGGGGCTACGAGCCCAACGAGATGCGACAGCTGCGTG----- 1940  
Db 381 LysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGlnValAla 400  
QY 1941 -----GCTGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976  
Db 401 PheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 415

RESULT 13  
AAB92693  
ID AAB92693 standard; prctein; 418 AA.  
XX  
AC AAB92693;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11083.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 93JP-00248036.  
PR 27-AUG-1999; 93JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 11083; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

XX Sequence 418 AA;

SQ Alignment Scores:

Pred. No.: 8 99e-133 Length: 418  
 Score: 1967.50 Matches: 385  
 Percent Similarity: 89.81% Conservative: 3  
 Best Local Similarity: 89.12% Mismatches: 10  
 Query Match: 48.51% Indels: 34  
 DB: 4 Gaps: 2

US-09-914-958B-35 (1-2177) x AAB92693 (1-418)

QY 753 ATGAGCCCTGAGCCAGTGAACCTCTCCCGGGGCAATCTCGCGCTCAGAGATGGCTGCT 812  
 Db 1 MetAspProGluAlaSerAspLeuSerArgGlyLeuAlaLeuArgAspValAlaAla 20  
 QY 813 GAGGCTGGGTCAAGAAATATACAGGGGGCCGAGGCTCAGCGTGGGCCCATCTCAGC 872  
 Db 21 GluAlaGlyValSerLysThrGlyArgGlyValSerValGlyProIleLeuSer 40  
 QY 873 AGCAGTGCCTCGGATATCTTCGCAATGAAATGGGCTAACTCTCTTTCCACAAC 932  
 Db 41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60  
 QY 933 CGGGGGATGGCACCTTTGTGACGCTGGCGGCGAGTCTGTGTGGACGACCCCAACAC 992  
 Db 61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80  
 QY 993 CATGGCGAGGTGTCGCCCTGCTGACTTCAACCGTGATGGCAAGTGCATCTCTAT 1052  
 Db 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100  
 QY 1053 GGCACTGGAATGGCCCCCACCAGCTCTATCTGCAATGAGCACCCATGGAGGTCCGC 1112  
 Db 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120  
 QY 1113 TTCGGGACATCGCTCAACCAAGTCTCCATGCCCTCCCTCTCGCACGGTCTATCACC 1172  
 Db 121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140  
 QY 1173 GCGGACTTGTACATGACAGGCTGGAGATCTTCTTCAACACATTCCTACCGCAGC 1232  
 Db 141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160  
 QY 1233 TCCTCAGCCACCCGCTCTTCCGCGCTCATCCGTAGAGACGAGACCCCTCATCGAG 1292  
 Db 161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180  
 QY 1293 GAGCTCAATCCGGGACGCTTGAGGCTGAGGCGCGGGCACAGGGGTGTGTGACC 1352  
 Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200  
 QY 1353 GACTTCGAGGAGAGGAGTGTGGACCTCATCTTGTCCATGGAGAGTCCATGGCTCAG 1412  
 Db 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220  
 QY 1413 CCGCTGTCTCTTCCGGGGCAATCAGGCTTCAACAACTGGCTGCGAGTGGTGCAC 1472  
 Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValPro 240  
 QY 1473 CCACCCGGTTTGGGCTTTTCCAGGGGAGCTAAGTCTGCTCTACACCAAGAGAGT 1532  
 Db 241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysSer 260  
 QY 1533 GGGGCCCCCTGAGGATCATCGAGGGGGCTCAGGCTACTGTGTGAGATGGAGCCGCTG 1592

Db 261 GlyAlaHisLeuArgIleLeuAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280  
 QY 1593 GCACACTTTGGCTGGGAGGATGAGCCAGCAGTGTGGAGGTGAGCTGCCAGATGCC 1652  
 Db 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300  
 QY 1653 AAGATGTTGAGCGGAACGTGGCCAGCGGGAGATGAACCTCAGTGTGGAGATCTCTAC 1712  
 Db 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320  
 QY 1713 CCCCGGATGAGACACACTTCAGGACCCAGCCCCCTGGAGTGTGGCCAGGATCTCC 1772  
 Db 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu----- 334  
 QY 1773 CAGCAGAAATGGCCATTCATGGACACCAATGAATGATCCAGTCCCATTCGCTGTC 1832  
 Db 335 -----ThrProMetAsnAlaSerSerHisSerCysAl 346  
 QY 1833 CCTCGAGACAAGCCCGTATGTCTCAACACTATGGAAGCTACAGGTCCCGGACCAACAAG 1892  
 Db 346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366  
 QY 1893 AAGTGCAGTCCGGGCTACAGCCCAACGAGGATGGCACAGCTGCTGGGCTGGTGAGC 1952  
 Db 366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly----- 383  
 QY 1953 CTGTGTTCAAGATAGTACACCAAGTTGGGAGAGCCTTGGTCCCTCAATCACTGAA 2012  
 Db 384 -----LeuSerAlaSe 387  
 QY 2013 TCACTGCCTTGAATCACCGCTGGAATACCTGTT 2046  
 Db 387 rHisArgAlaProAlaProProProProProLeu 398  
 RESULT 14  
 ABR48170  
 ID ABR48170 standard; protein; 418 AA.  
 XX ABR48170;  
 AC ABR48170;  
 XX  
 DT 12-JUN-2003 (first entry)  
 XX  
 DE Human bladder cancer associated protein sequence SEQ ID NO:56.  
 XX  
 KW Human; bladder cancer; cytostatic; gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003003906-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 03-JUL-2002; 2002WO-US021339.  
 XX  
 PR 03-JUL-2001; 2001US-0302814P.  
 PR 03-AUG-2001; 2001US-0310099P.  
 PR 08-NOV-2001; 2001US-0343705P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX  
 PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 PI Mack DH, Aziz N;  
 XX  
 XX WPI; 2003-201532/19.  
 DR N-PSDB; ACC50981.  
 XX  
 PT Detecting a bladder cancer-associated transcript in a cell from a  
 PT patient, comprises contacting a biological sample from the patient with a  
 PT bladder cancer-associated polynucleotide or antibody.  
 XX  
 PS Claim 10; Page 248; 307pp; English.

XX The present invention describes a method for detecting a bladder cancer-  
 CC associated transcript in a cell from a patient. The method comprises  
 CC contacting a biological sample from the patient with a polynucleotide  
 CC that selectively hybridises to a sequence that is 80 % identical to a  
 CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059  
 CC encode the human bladder cancer-associated proteins given in ABR48146 to  
 CC ABR48442). Bladder cancer-associated sequences from the present invention  
 CC have cytostatic activities, and can be used in antisense gene therapy and  
 CC in vaccine production. The method can be used for detecting a bladder  
 CC cancer-associated transcript in a cell from a patient. The method is  
 CC useful in diagnosing or treating bladder cancer and in screening for  
 CC compounds that modulate bladder cancer, such as hormones or antibodies.  
 CC The nucleic acid molecules from the present invention may be used in  
 CC various screening and diagnostic methods, and for gene therapy, vaccine  
 CC and/or antisense/inhibition applications  
 XX

SQ Sequence 418 AA;

#### Alignment Scores:

Pred. No.: 8.99e-133 Length: 418  
 Score: 1967.50 Matches: 385  
 Percent Similarity: 89.81% Conservative: 3  
 Best Local Similarity: 89.12% Mismatches: 10  
 Query Match: 48.51% Indels: 34  
 DB: 6 Gaps: 2

US-09-914-958B-35 (1-2177) x ABR48170 (1-418)

QY 753 ATGGACCTGAGGCGCAGTACCTCTCCCGGGGCACTCTGGCGCTCAGAGATGTGGCTGCT 812  
 Db 1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAla 20  
 QY 813 GAGGTGGGTGAGCAAAATATACAGGGCGAGCGGTGAGCGTGGGCCCATCTCAGC 872  
 Db 21 GluAlaGlyValSerLysTyrThrGlyArgGlyValSerValGlyProIleLeuSer 40  
 QY 873 AGCAGTCTGGATATCTTTCGCACATGAGTGGCGCTAACTCTCTTTCCACAAC 932  
 Db 41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60  
 QY 933 CGGGCGGATGGACCTTTGTGACCTCGGCGCAGTGTGGTGGAGCAGCCCCCACCAG 992  
 Db 61 ArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspProHisGln 80  
 QY 993 CATGGCGAGGTGTCGCTGCTGACCTTCAACCTGATGGCAAGTGGACATCTCTAT 1052  
 Db 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100  
 QY 1053 GGCACCTGGAATGGCCCCCAGCCCTCTATCTGCAATGAGCACCATGGGAAGTCCGC 1112  
 Db 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120  
 QY 1113 TTCGGGACATCGCTCACCCCAAGTTCTCCATGCCCTCCCTGTCGCGACGGTCAATACC 1172  
 Db 121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140  
 QY 1173 GCCGACTTTCGAATGACGAGAGCTGGAGATCTTCTTCAACAAATGCTACCGCAGC 1232  
 Db 141 AlaAspPheAspAsnAspGlnGluIlePhePheAsnAsnIleAlaTyrArgSer 160  
 QY 1233 TCCTCAGCCACCGCTCTTCGCGCTCATCTCCGTAGAGACGAGACCCCTCATCTGAG 1292  
 Db 161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180  
 QY 1293 GAGCTCAATCCCGGCGAGCCCTTGGAGCTTGAGGCGCGGGGCACAGGGGTGTGTGACC 1352  
 Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200  
 QY 1353 GACTTCGAGGAGACGGGATGTCGACCTCATCTTGTCCCATGGAGATCCATGGCTCAG 1412  
 Db 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

QY 1413 CCCTGTCCGCTCTCCCGGGCAATCAGGGCTTCAACAACAACTGCTGCGAGTGTGCGCA 1472  
 Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240  
 QY 1473 CGCACCCCGGTGTGGGCTTTTCCAGGGAGCTAAGGTCTGCTCTACACCAAGAGAGT 1532  
 Db 241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260  
 QY 1533 GGGGCCCCACCTGAGATCATCGAGGGGCTCAGCTACCTGTGTGAGATGGAGCCCGTG 1592  
 Db 261 GlyAlaHisLeuArgIleIleAspGlySerGlyTyrLeuCysGluMetGluProVal 280  
 QY 1593 GCACACTTTGGCTCGGGAGAGATGAGCCAGCAGTGTGGAGGTGACGTGCCAGATGCG 1652  
 Db 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300  
 QY 1653 AAGATGTGTAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTCTGTGAGATCTCTAC 1712  
 Db 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320  
 QY 1713 CCGCGGATGAGGACACACTTTCAGACCCCGCCCTGAGTGTGGTGGCCAGGATCTCTCC 1772  
 Db 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu----- 334  
 QY 1773 CAGCAGGAAAATGGCCATTGTCATGGACACCAACCAATGAATGCATCCAGTTCCTCATTCGTGTC 1832  
 Db 335 -----ThrProMetAsnAlaSerSerHisSerCysAl 346  
 QY 1833 CCTCGAGACAAGCCGTTATGTCTACACCTATATGGAAGCTACAGTGGCGGACCAACAG 1892  
 Db 346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366  
 QY 1893 AAGTGCAGTCCGGCTACGAGCCCAACGAGATGTCACAGCTGGTGGTGGTGGTGGAGC 1952  
 Db 366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly----- 383  
 QY 1953 CTGTGTTGAAGATAGTCACACCAACAGTGGGAGAGCCTTGGTCCCTGAATCACTGAA 2012  
 Db 384 -----LeuSerAlaSe 387  
 QY 2013 TCATGCTTGAATCACCCTCGGATACCTGTT 2046  
 Db 387 rHisArgAlaProAlaProProProProLeu 398  
 RESULT 15  
 ABUS6635  
 ID ABUS6635 standard; protein; 418 AA.  
 AC ABUS6635;  
 DT 02-APR-2003 (first entry)  
 XX Lung cancer-associated polypeptide #228.  
 DE Lung cancer-associated polypeptide #228.  
 KW Lung cancer-associated polypeptide; cytostatic; emphysema;  
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 XX Unidentified.  
 OS  
 XX  
 PN WO200286443-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 XX 18-APR-2002; 2002WO-US012476.  
 PR 18-APR-2001; 2001US-0284770P.  
 PR 10-MAY-2001; 2001US-0290492P.  
 PR 09-NOV-2001; 2001US-0329245P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-0334370P.





This Page Blank (uspx)



QY 1716 ---GGGGTAGAGATCTCCAGCACTGAGTTCATCTCCCGCTGGCCAGCTTCGGCTCA 1660  
 Db |||||  
 QY 121 AlaSerArgArgSerPheProAlaLeuMetMetArgProCysTrpProLeu----- 137  
 QY 1659 CCATCTTCCCATCTGGCCACCTCACTCCACACTGCTGGCTTCATCTTCCCGAGGCCAA 1600  
 Db |||||  
 QY 138 ---SerMetArgLeuAlaSerSerLysProCysTrpLeuGluLeuSerProTrp 156  
 QY 1599 AGTGTGCCACGGCTCCATCTCACACAGGTAGCTGAGCCCGCTCGATGATCCTCAGGT 1540  
 Db |||||  
 QY 157 ArgLeuSerArgTrpPheSerArgSerMetLeuSerGlyProArg----- 171  
 QY 1539 GGGCCCACTCTTCTGTGTGAGACACGACTTAGCTTACCTCCCTGGCA-----AAGGCC 1487  
 Db |||||  
 QY 172 -----AlaLeuLeuArgPro-----ProTrpLeuSerArgSerPro 183  
 QY 1486 CCAACCGGTGCGTGGCACCACCTCGCAGCCAGTGTGTGTTGAACCCCTGATGCCCCGG 1427  
 Db |||||  
 QY 184 ProArgArgSerArgLeuTrpProArgSer----- 193  
 QY 1426 AAGACGGACAGGGCTGAGCCATGACTCTCCATGGGACAGATGAGTCCAGCATC--- 1370  
 Db |||||  
 QY 1369 -----ProThrSerArgProTrpArgCysThrArgArgSerThrVal 207  
 QY 208 ArgArgGserProArgAlaGluLeuSerProProSerAlaGluSerThrArgProAla 227  
 Db |||||  
 QY 1321 GGCTCCAAAGGCTCCCGGATGAGTCC-----TCG 1289  
 Db |||||  
 QY 228 ValSerSerSerSerSerLeuValSerArgProProArgLeuSerArgLeuAla 247  
 QY 1288 ATGAGGGGTCTCCGTCTCTACGGATGACCGGAGAGCGGTGGCTGAGGAGCTG 1229  
 Db |||||  
 QY 248 LeuArgArgSerAspCys---LeuProLeuArgAlaProProArgLeuSer----- 263  
 QY 1228 CGTGGGATGTTGTTGAGAGATCTCCAGCTCTCTG----- 1190  
 Db |||||  
 QY 264 ArgSerAlaLeuLeuThrSer-----SerThrProTrpLeuLeuAsnTrpProProSer 281  
 QY 1189 TCATTGTCAAGTCCGGCTGATACCGTGG----- 1157  
 Db |||||  
 QY 282 AlaLeuSerSerAlaProValAlaThrLeuArgProProAlaLeuProGluSerArgProLeu 301  
 QY 1156 -----ACAGGGAGGCGCATGGAGAAC----- 1136  
 Db |||||  
 QY 302 ProLeuLeuSerArgGluAlaLeuAlaSerValArgAlaProGluThrValProSer 321  
 QY 1135 -----TTGGGTGAGCGATGTCGCGGAGCGGACCTTCCCA----- 1100  
 Db |||||  
 QY 322 AlaTrpLeuSerSerAlaProSerThrArgArgSerMetProAlaArgLeuAlaArgVal 341  
 QY 1100 ----- 1100  
 Db |||||  
 QY 342 ProLeuAlaLeuLeuSerArgArgAlaAlaAlaMetSerArgProAlaAlaProGluLeu 361  
 QY 1099 -----TGCGTCTCATTTGCATAGAGAGCGGTGGGG 1067  
 Db |||||  
 QY 362 ThrProAlaArgLeuLeuSerArgAlaTrpAlaMet---ArgArgValArgLeuAlaAla 380  
 QY 1066 CCATTCCAGTTGCATAGACGATGCTCCACTTCCCATCGGTGAAGTCAGCCAGGGG 1007  
 Db |||||  
 QY 381 ProileArgArgProProArgLeuLeuSerArgLeuAlaAlaAlaArgValLysAlaArgSer 400  
 QY 1006 -----ACACTTCGCGCATGCTGGTGGGGGTGCTCCACA 974  
 Db |||||  
 QY 401 LeuGluLeuSerProProAlaProLeuSerThrAlaProSerTrpArgSerSerAla 420  
 QY 973 CCAGCACTGGCC----- 962  
 Db |||||  
 QY 421 ProGlyValAlaAlaAlaArgProProSerArgLeuThrSerValProProLeuArgSerArg 440  
 QY 961 -----GCAGGTCCCAAAGGTCCCATCGCCCGGTTGTGGAAAGGAAGTTAGGC 911

Db 441 ValThrProValSerLeuThrArgThrProProArg---TrpSerArg----- 456  
 QY |||||  
 QY 910 CCATCTCATTTGTCGCAAG-----ATATCC 884  
 Db |||||  
 QY 457 PropheArgLeuAlaSerSerArgProLeuAlaLeuThrArgProCysTrpProLeuSer 476  
 QY 883 GAGCAGCTGCTGAGGATGGGCGCCAGCTGAGCGCTCGGCCCTGTATATTGTGTG 824  
 Db |||||  
 QY 477 SerArgAlaAlaThrArgValMetProAlaLeuLeuProMetProProLeuArgLeuLeu 496  
 QY 823 -----ACCCAGCTCAGCAGCCCATCTCTGAGCGCCAGCATGCCCGGGAGAGG--- 773  
 Db |||||  
 QY 497 SerSerThrProValArgIleSerThrAlaProLeuAlaLeuThrProGluArgLeu 516  
 QY 772 -----TCATGGCTCAGGTCATTTCAATGAGGCGATCAGGCGCCACATTCOGTAG 719  
 Db |||||  
 QY 517 LeuSerArgLeuAlaProCysSerValThr---ProAlaSerLeuSerSerLeuPro--- 534  
 QY 718 GCGTAATTGGCAATGTAGATAGAGTAGCGTCCAGAGCCCTTCTCTCCACACAGAGCC--- 662  
 Db |||||  
 QY 535 -----ProTrpLeuSerSerAlaAlaTrp 542  
 QY 661 -----ACAGAGCTCCGCAAGAGGCTGGCCACACAGCGGCGCAGCTTGACCTCA 611  
 Db |||||  
 QY 543 LeuValSerAlaSerAlaProAlaLeuGluLysValProProArgLeu----- 558  
 QY 610 TCGCTCAGAGATGCTTCCACCGGTTATGCGGAACCTTGAACAACTTTCGGGTGACGTG 551  
 Db |||||  
 QY 559 SerArgArgAlaAlaArgAlaValArgLeuProSerLeuThrSerAlaProProTrpLeu 578  
 QY 550 GCCACCCCGAGAGGCATTATTGTTGTTGAGGAAGTAGATCTCTCCCGCGCGTCCCGC 491  
 Db |||||  
 QY 579 SerSerThrProProArgLeuThrLeuArgLeuPheTrpLeuSerSerArgPro---Pro 597  
 QY 490 TCGATGTCGAGCTGTGACCCCAATGGCGTTCCCTCGCGTCCGCGAGCGGTAGTAG 431  
 Db |||||  
 QY 598 SerProLeuAsnSerSerProPro----- 605  
 QY 430 GGTGAGCTGCGCTCATCGACCGGATGTTACACAGCGCTTCTGGGCGCGGTCACTTTC 371  
 Db |||||  
 QY 606 SerArgLeuArgProSerArgProAlaSerThrProLeuAlaTrpPheSerArgCys 625  
 QY 370 AGAACCCAGGTTCGGTCCATTGTACCCCGCCAGCAGATCTCAAAGTCCCATCATGTCC 311  
 Db |||||  
 QY 625 ThrValArgArgProLeuSerProMetThrLeuProProArgLeuSerSerCysSer 645  
 QY 310 ACATCAGTAACCTGCCACA----- 293  
 Db |||||  
 QY 646 ArgAlaPheThrAlaThrPheGluValLeuGluIleSerProAlaArgLeuSerThrCys 665  
 QY 292 -----CCATAGTTGAGCTGGTGGGATTACTGTCTCATAG 260  
 Db |||||  
 QY 666 ArgAlaSerIleAlaMetProPheAlaAlaIleSerGlnAspTrpLeuLeuSerIle 685  
 QY 259 TCAGGAGGCGAGCACTGAGTTGGTACTGACGTGAGTGAACATCGGTTTCAGCCCGCTGGGAC--- 203  
 Db |||||  
 QY 686 ValSerAlaAlaThrPheArgValPheSerLeuThrSerSerProCysTrpAlaArg 705  
 QY 202 ---CCCTCAGTGTGGCGAGAAC----- 182  
 Db |||||  
 QY 706 LeuProSerValAlaCysArgSerProLeuAlaAlaIleArgProProAlaLeuProThr 725  
 QY 181 -----CAGAGCAGCAGGAGCGGTAACTCTCGACATCGCGGGTCAGCGCTCGGA 128  
 Db |||||  
 QY 726 AlaPheAlaAlaArgSerArgProSerLeuAsnSerLeuProProTrpSerLeuArg 745  
 QY 127 GCCATC-----CTCCGCTCTCGGCGCCCGCTAGGGG 92  
 Db |||||  
 QY 746 ArgLeuAlaThrSerThrArgThrProAlaLeuProLeuAlaLeuProProSerArg--- 764  
 QY 91 GTGGGAAGCGGGCGCTCGCTGCGCGCTTCCCGCGCGCGCGCTGCTGCTCCAG 32

```
Db 765 -----LeuSerLysAlaProProValAlaLeuLysProAlaSerAlaThrSerThr 781
QY 31 CCCCGTCCCGGCTGGCTCGAGCCT 5
Db 782 ProProArgLeuSerArgLeuAlaPro 790

RESULT 2
US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31760
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31760

Alignment Scores:
Pred. No.: 5,24e-10 Length: 1706
Score: 240.00 Matches: 190
Percent Similarity: 29.12% Conservative: 36
Best Local Similarity: 24.48% Mismatches: 219
Query Match: 5.81% Indels: 331
DB: 4 Gaps: 40

US-09-914-958b-35 (1-2177) x US-09-252-991A-31760 (1-1706)
QY 1779 CTGCTGGGAGATCTTGGCCACATCCAGTGGGGTGGTCTGAAGTGTCTCTCAT 1720
Db 126 ProAlaAspArgLeuProGluAlaArgArgProGlySerGlyAlaGlnAla----- 142
QY 1719 CCGGGGGTAGAGGA---TCTCCAGACACTGAGTTCATCTCCCGC-----TGGCCA 1672
Db 143 ProGlyGlyArgGlyProGlyProAlaHisArgHisAlaGlnArgArgGlnProArgPro 162
QY 1671 CGTTCGGCTCACCATTGTCCTGTCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCT 1612
Db 163 ArgThrGlyAspProGln-----HisPro 170
QY 1611 -----TCCCGAGGCCAAAGTGGCCA--- 1591
Db 171 ArgAlaAlaArgSerAlaValArgArgArgProGlyLeuGlyAlaProAla 190
QY 1590 CGGGTCCATCT-----CACACAGTAGCTGAGCCCGCT 1555
Db 191 ArgAlaProGlyGlyAlaTyArgGlyAlaGlyArgArgThrGlyProLeuProGlu 210
QY 1554 CGATGATCTCGAGTGGGGCCCACTCTTCTGTGTAGACGACGCTTAGCTTCC----- 1500
Db 211 ProAlaProArgGlyHis-----ArgAla-GluLeuProAlaProAl 226
QY 1499 -----CTGGCAAGGCC 1486
Db 226 aArgThrAspArgAlaGlyGluArgLeuAlaAlaGlyLeuProGlyProAlaPr 246
QY 1485 CAAACCGGTGCGTGCACCACTCGCAGCCAGTTGTTGTAAGCCCTGATGCCCGGA 1426
Db 246 oAlaProGlyAlaGlyHisArgGlnPro----- 256
QY 1425 AGACGACAGCGGTGAGCCATCGACTCTCCATGGGACAAGATGAGTCCAGTCCCGCT 1366
Db 257 -----Gly-AlaProGlyGlyLysLeuArgAlaAlaGlyPro- 268
```

```
QY 1365 CTCGTCGAAGTCGTCACACACCCCTGTGCCCCG---CCCTCAGSCTCCAGGCG- 1310
Db 269 -----ProLeuAlaThrGlyProLeuArgArgGlnProAlaGlyAlaThrAlaA 285
QY 1309 -----TCGCGGGATTGAGCTCTCGATGAGGGGTCTCCGTCCTCTTACGGA 1261
Db 285 rgLeuArgHisArgProGlyHisProAlaArgLeuArgGlnProAlaArgLeuArgA 305
QY 1260 TGACGCGGAAGAG----- 1247
Db 305 rgSerGlyGluArgProAlaGluSerProArgArgThrProGlyGluLeuArgGlyArgA 325
QY 1246 -----CGGTTGGCTGAGGAGCTGCGTAGG-----CAATGTTGT 1213
Db 325 rgProAspAlaLeuProArgLeuAlaGlyLys-ProGlyArgTyProAlaGluPheAla 344
QY 1212 TGAAGAAGATCTCCAGCTCTCTGTT-----CATTTGCAA 1180
Db 345 ArgGlyAlaAlaProAlaProGlyProLeuAlaAlaValProArgArgHisAlaArg 364
QY 1179 AGT-----CGGCGGTGATGACCGTCGGACAGGGGCGGCGCATGGAGACTGG 1132
Db 365 AlaAlaGluProTrpArgArg-----ProAlaGlyGlyArgAlaAspArgThr--- 381
QY 1131 GTGAGCGATGTCCCGAAGCGACCTTCCCATGGTGTCTATTGTCAGATAGAGCGGT 1072
Db 382 -----AlaThrAlaProAlaArgProArgThrAlaCysAlaAlaVal 396
QY 1071 GGGGGCATTCAGTTGCCATGACGATGCCATTCGTCATCCGTTGAAGTCAGCCA 1012
Db 397 AlaGlyProArgHisAlaArgArgGluGlnProArgThrLeuProGlyProArgProPro 416
QY 1011 GGGCGCACCTCCCGCATGCTGCTGG----- 985
Db 417 GlyAlaArgGlyThrHisLeuAlaGlySerAlaGlnProAlaAlaProAlaAlaArg 436
QY 984 -----GGTCTCCACACACGACACTGGCCCGCAG 958
Db 437 ProProAlaAlaLeuProAlaArgAlaArgArgGlyArgArgGlnPro---ProGly 455
QY 957 GTTCCCAAAGGTGCCATCGCCCGGTTGT----- 928
Db 456 ArgProProArgArgLeuArgProGlyAlaProValAlaAlaAlaProAlaThrGly 475
QY 927 -----GSAAGAAGTATTAGCCCATTCATTCGTCGCAAGATAT 886
Db 476 GlyAlaAlaThrGluProGlyAlaGlyAspSerArgHis---ArgCysArgSerGlyThr 494
QY 885 CCGAGSACTGTGCTGAGGATGGGGCCACGCTGACGCTCGGCCCTGTATATTTC 826
Db 495 AspArgPro----- 497
QY 825 TGACCCCGCTCAGGACCACTCTCTGAGCGCCAGAAATGCCCGGAGAGTCACTCG 766
Db 498 ---GlyArgProArgGlnProAla-----AlaAlaArgAspProGlyProGlySerTtp 514
QY 765 CCTCAGGGTCCATTTCAATGAGGCGCATCAGGGCCACATTACCGTAGGCGTAATGGCAA 706
Db 515 Pro-----GlyProSerAlaAlaArg----- 522
QY 705 TGTAGATAGTAGTGGCTCCAGAGCCCTTCTCTCCACACAGGCCACAGAGCTCCGCAA 646
Db 523 -----AlaAlaArg 525
QY 645 AGAGCTGGCCACACACAGGGCCAGTTGACCTCATCGCTCAGGATGCTTCCCGCGGT 586
Db 526 ArgGlyAspProHisArgArgGlu----- 533
QY 585 TATTCGGGAAGTGTGAACAACTTGTGCTGTAGTGGCCACCCCGAGAGGCGATTATTGG 526
Db 534 -----ThrAlaLeuArgSerAlaAlaArgArgProArgArg----- 546
```

```
QY 525 TGTGAGGAAGTAGATCTCTCCCGCGCT----- 496
Db 547 -----AlaAspProProGlyArgThrProAlaGlnProGlyGlnThrGluGly 562
QY 495 -----CCCGTCGATGCGCAGG 478
Db 563 AlaArgArgLeuAlaArgArgArgThrAlaGlyThrAlaAlaProArgHisArgArg--- 581
QY 477 CTGTGACCCCAATGGCGTCCCGTCCCGTCCC-----GCAGCG 439
Db 582 -----GlnProGlnProAlaGlyProArgProProAlaArgGlyAlaThrAla 598
QY 438 CGTAGTAGGTGAGTGC-----GCTCATCGACCGCA 406
Db 599 ThrSerArgLeuProAlaLeuProTrpProArgArgProAlaArgSerHisArgProGly 618
QY 405 TGT---TCACAGCGCGTCTTGGCGCGGTTCATCTCAGAACCCAGGTGGTTCATTGT 349
Db 619 IleProGlyProAlaAlaProAlaProGlyThrThrArgArgAspSer----- 634
QY 348 ACCCGCCACGACGATCTCAAGTCCCATCATGTCACATCACTGACCAACCAT 289
Db 635 -----ArgArgArgThrArgSerHisGlyAlaProHis-----AlaGlnProArg 649
QY 288 AGTTAGCTGGTGGATTACTGTCTATGTCAGGAGGAGACTGAGTTGGTGAATGCAG 229
Db 650 SerAlaAlaGly-----LeuGluAlaAlaProGlyThrArgGlnSer 663
QY 228 TGAACATGGTTTCAGCGCTGGGACCTTCAGTGTGGTGGAGAACCA----- 180
Db 664 GlyThrAlaValGlnProGlyArgSerGln-GlyGlnGlyArgGlnProArgArgSerAr 683
QY 179 -----GAGCAGCAGCAGGAAACGGTAACATCC 154
Db 683 GCysAlaAspAlaArgAlaGluProArgThrAlaArgGlnGlyHisArgArgAlaPr 703
QY 153 TGGACATGCGGGTTCAGCGCTCGGAGCCATCTCCC-----GCTCTGGCGCCGC 103
Db 703 OGlyGlnProGlnAlaMetGlyArgHisProProAspArgGlnAlaLeu-AlaAlaP 723
QY 102 CCCTTAGGGCTGGGAAGCGCGCTCGCTGCGCTCTGCGCGCGCGCGCG----- 50
Db 723 roAlaArgValHisArgAlaGlyArgArgThrArgProAspGluProAlaProGlyLeuA 743
QY 49 -----CCGCTGTCTCTCCAGCCCGCGTCCCGGCT 17
Db 743 laAspGluProArgProGlyGlnProArgAlaAlaAla 755

RESULT 3
US-09-252-991A-18296
; Sequence 18296, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18296
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18296

Alignment Scores: 3.29e-10 Length: 467
Pred. No.: 239.50 Matches: 160
Score:
```

```
Percent Similarity: 32.81% Conservative: 27
Best Local Similarity: 28.07% Mismatches: 185
Query Match: 5.79% Indels: 199
DB: 4 Gaps: 40

US-09-914-958B-35 (1-2177) x US-09-252-991A-18296 (1-467)

QY 1815 GGATGATTCATTGGTGTCCATGCAATGGCCATTTTCTGTGGGAGATCTCTGGCCAC 1756
Db 6 GlyCysProAlaGlyCysGlyArgGlyArgGlyAsnTrpProAlaArgAlaGlyPhe 25
QY 1755 ACTCGAGTGGGCTGGTCTGCTGAAGTGTCTCATCCGGGGGTAGAGGATCTCCAGCA 1696
Db 26 ProPro-----ProProAlaCys---ArgProGly----- 34
QY 1695 CTGATTCATCTCCCGCTGGCCAGTTCGGC---TCACATCTTGGCATCTGGCCAG 1639
Db 35 ---ThrGlySerSerArgTrpAsnArgAlaGlyArgAlaArgGlyCysProGlyCysAla 53
QY 1638 TCACCTCCACACTGTGCTTTCATCTTCCCGAGCCAAAGTGTCCA----- 1591
Db 54 AlaProProAlaGly-----ValProGlyAlaGlyArgSerValProAlaProAlaCys 71
QY 1590 -----CGGCTCCATCTCACACAGTAGCTGAGCCCGCTGATGATCC 1546
Db 72 ArgArgArgHisArgArgSerArgProSerSerGlyArgProArgGlnGlyArg 91
QY 1545 TCAGTGGGCGC-----CACTCT 1528
Db 92 ArgGlySerProThrAlaAlaLeuArgLeuArgProAlaArgArgArgHisPro 111
QY 1527 TCTTGTGTGAGA-----GCACGACCTTAGTCTCCCTGGCAAGGCCCAACCGGG 1477
Db 112 AlaTrpArgArgSerAlaSerArgPro-----AlaGlyAlaProProAlaPro 129
QY 1476 TCGTGGCACCCTCGACGCCAGTGT-----TGTGA 1444
Db 130 CysProAlaProArgGlyAlaGlyAlaProSerValArgGluArgProAlaGlySerAla 149
QY 1443 AGCCCTGATTGCGCGAAGACGGACAGCGGTGACCCATGACCTCTCCATGGGCAAGA 1384
Db 150 ProProProArgProAlaArgArgProProAlaProPro-----HisArgThrGly 166
QY 1383 TGAGTCTCAGCATCCCGTCTCCGTCGAAGTCGTCCACACACCCCTGTGCCCGGCCCT 1324
Db 167 ---GlySerGlyTrpArgArgArgArgGlyArgThrArgPro----- 180
QY 1323 CAGGCTCAAGGCTGCGCGGATTCAGCTCCT-----CGATGAGGGGTCTCCGTGCT 1270
Db 181 ---AlaAlaSerArgArgProAlaGlyGlyProAlaGlyArgProGlyGlySerArgAsn 199
QY 1269 CTCTACGGATGACGCGGAAGAGCGGTGGCTGAGGAGCTGC-----GATGGCAATGT 1216
Db 200 -----ArgArgArgAlaGly-----CysArgProAlaArgProVal 211
QY 1215 TGTGTGAAGAAGATCTCCAGCTCTCTGTCATTTGCAAGTCGGCGGTGATACCGTGGCA 1156
Db 212 AlaArgAlaGlyProGlyProGlyProGlyArgCysArgArgArg-----Cys--- 227
QY 1155 CAGGGAGGGCATGGAGAACTTGGTGGCGGATGTCGCGAAGCGGACCTTCCCATGG 1096
Db 228 -----AlaTrpSerPro-----CysPro----- 233
QY 1095 TGTCTATTTCAGATAGAGCGGTGGGGGCCATTCCAGTTGGCCATAGACATGTCACCT 1036
Db 234 ---ThrProArgAspArgAlaGlyAlaGly-----GCCATGCTGGT 988
QY 1035 TGCCATCAACGGTTGAAGTCAGCCAGGCGCACCTC-----AspArgHisValProAlaGlyAspAlaAla 254
Db 243 ----- 937
QY 987 GGGGTGCTCCACACACAGCACTGCGCAGCGTCCCAAAAGGTGCCATCG----- 937
```

```

Db      255 Gly-----TtrPro---ThrGlySerArgCysAlaArgArgSerGly 267
QY      936 ---CCCGCTTGTGAAAGGAGTTAGCCCATTCATTTGCGCAGAGATATCCGAGG 880
Db      268 SerAlaGlyCysAla-----ArgProProArg 276
QY      879 CACTGCTGCTGAGATGGGCCCGCCAGCTG-----ACGCCT 845
Db      277 ArgSerCysValArgTtrAlaAlaArgValAlaArgAspSerAlaArgArgSer 296
QY      844 CGGCCCTGTATATTGCTGACCCCGCCAGCCAGCCACATCTCTG---AGGCCAG 788
Db      297 ArgPro-----ValArgSerAlaLeuArgGlyLeuProAlaAlaArg 310
QY      787 ATGCCCGGAGAGTCACTGGCTCAGGCTCCATTTCAATGAGGGCATCAGGCCACACA 728
Db      311 ArgSerArgProArgHisArgArgProGlySerValProValArgThrSerArgAla 330
QY      727 TTACCGTAGGCGT-----AATGGCAATGT-----AGATAGATAGGCTCCA 686
Db      331 ThrPro-ArgArgTtrHisAlaArgTtrGlnAlaAlaAlaArgArgProGlySerVal 350
QY      685 GAGCCCTTCTGTCCACAGCCGACAGAGCGTCCGGCAAGAGGCTGGCCACACCCAGG 626
Db      350 userSerAlaThrProAlaAlaPro-----GlyValArgGlyTyrProArgGluTyr 367
QY      625 GCCACGTTGACCTCATCGCTCAGGATGCTCTCCACCGGTTATTGCGGAATCGAAACAC 566
Db      367 rAlaArgAlaProArgProAla-----LeuPro----- 376
QY      565 TTGCTGCTGTACGTGGCCACCCGAGAGGCATATTGTTGTTGAGGAAGTAGATCTCC 506
Db      377 -----GlyArgProAlaAlaArg----- 382
QY      505 TCCCGCGCTCCCGCTGATGTGCGAGCTGTGACCCCAATGGCGTTCCCGCTGCC----- 451
Db      383 -ProAlaArgProArgProCysArgArg-----ProAlaAspArgSerProGlyGlnG 400
QY      450 -----GGTCCCGCAGCGGTAGTAGGTGAGCTGCGCTCATCGACCGCATGTTTC 401
Db      400 yArgProProGlySerSerAlaThrArgArgTtrArgProCysHisArg--ArgAlaG 419
QY      400 ACCAGCGCTTCTGGCCCGGTCTATCTCAGAACCCAGGTTGGTCTCATCTGTACCCGCC 341
Db      419 uProAlaAlaProAlaGlyTyr-----ProGlyAlaAlaArgSerAlaTh 436
QY      340 ACAGCATCTCAAGTCCCATCATGCT 313
Db      436 xArgArgThrHisAlaGlyArgArgGly 445

```

## RESULT 4

```

US-09-252-991A-24567
; Sequence 24567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24567
; LENGTH: 681
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24567

```

## Alignment Scores:

```

Pred. No.: 1.02e-09 Length: 681
Score: 234.00 Matches: 194
Percent Similarity: 30.26% Conservative: 36
Best Local Similarity: 25.53% Mismatches: 242
Query Match: 5.66% Indels: 288
DB: 4 Gaps: 45

US-09-914-958B-35 (1-2177) x US-09-252-991A-24567 (1-681)

QY      1801 GTGTCATGCAATGCAATTTCTGCTGGAGAGATCTCTGGCCA---CACTCCAGTGGG 1745
Db      1 ValThrHisAlaTtrPro-----AsnTtrProProTtrAlaCysAsnSerThrGly 18
QY      1744 GCTGGGCTCTGAAGTGTCTCTCATCCGGGGGTAGAGATCTCCAGCAGCTAGTTCATC 1685
Db      19 -----ArgThr-ProThrAspGlyArgSerGlySerThrArgArgSerSerAl 34
QY      1684 TCCCGCTGGCCAGCTTCCGGCTCA----- 1660
Db      34 aAlaCysTtrPlyHisSerAlaIleArgArgValArgAsnArgSerAsnArgAlaTr 54
QY      1659 -----CCATCTTGGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCTTC 1610
Db      54 pProArgTtrProGlyCysAlaArgThrAlaProThrSerAspCysTtrSerAlaAsnAl 74
QY      1609 CCCAGGC-----CAAAAGTGTGCCCGGCTCCATCTCACACAGGTAGCTGAG 1562
Db      74 aAlaSerArgTtrSerIysArgSerAlaProProAlaProArgAlaSerTtrSerThrAr 94
QY      1561 CCCCGCTGATGATGCTCAGGTGGGCCCCACTCTTCTTGG-----TGTAG 1517
Db      94 gThrAlaArg-----GlySerProCysAlaTtrMetProIysAlaAlaCysAr 110
QY      1516 AGCAGCAGCTTAGCTCCCTCCGGAAGGCCCAACCGGGTGGTGGCCACCATCGCAGC 1457
Db      110 gArgArgPro----- 118
QY      1456 CAGTTGTTGTGAAGCCCTGAT-----TGCCCCGGAGACGACGCGCTGAGCC 1406
Db      118 gSerCysArgTtrPserSerGlyAsnGlyArgTtrArgTtrArgArgAlaValArgPr 138
QY      1405 ATGAGCTCTCCATGGACAAAGATGAGGTCCAGCATCC---CGTCTCCGTCGAAGTCGGTC 1349
Db      138 oTtr-----ProAlaSerProArgValAlaArgValAlaGl 150
QY      1348 A-----CCACACCCCTGTGCCCCCGCTCAGGCTCCAGGCTCGCGCGGATGAGC 1295
Db      150 yValTtrProProSerLeuCysProAlaPro-----ProArgArgArgArgProGlyAr 168
QY      1294 TCCTCGATGAGGGGGTCTCCGTCTCTACGGATGACGCGGAAGAGGC----- 1246
Db      168 gGlnArg-----ArgProArgGlyProAlaAlaGl 178
QY      1245 -----GGTTGGCTGAGGAGTGGCGTAGGCAATGTTGTGAAGAAGATCTCCAGC 1196
Db      178 nArgArgAlaAlaTtrArgArgArgAlaGlyAspGln-----ProAl 192
QY      1195 T-----CCTGCTCATTCGCAAGTCGGCGGTGATGACCGTGGGACACAGGG---AGGGC 1145
Db      192 aAlaArgProGlyArgGlyGlnArg-----ThrCysLeuGlnProLeuLeuAl 208
QY      1144 ATGGAGAACTTGGGTGAGCGCATGTCGCGGAAGCGGACCTTCCCATGGTGTCTATTGC 1085
Db      208 aValGlnProAlaValLeuGlnCysProAla-----ArgArgAlaGlyAspHisSe 225
QY      1084 AGATAGAGCGGTGGGCCCATTCAGTTGCCATAGACGATGTCCCATTTGCCATCACCG 1025
Db      225 rArgArgGlySerGlyGlyThrGlyAspProSerArgArgProGlyArgGlyAspGlyPr 245
QY      1024 TTGAAGTCAGCCAGCGGCCACCTCGCCCATCTGCTGGTGGGGTCTGCCACACACGCTG 965
Db      245 oProGlyGluProGlyAlaAspArgLeuAspSerGlyGlyProAlaHisAlaProAl 265

```





Db 182 rProProArgSerArgLeuThrProArgSer-----ProThrSerAr 197  
QY 439 -----CGTCGGGACGGCAGGGAAGCCATTGGGGTCACAGCTGGACATCGAC 491  
Db 197 gProTirpArgCys---ThrArgArgSerThrValArgArgSerProArgAlaGluI 216  
QY 492 GGGGACGGCCGGGAGGAGATCTACT-----TCCCAACACCAATAATGCC 536  
Db 216 eSerProSerAlaGluSerThrArgProAlaValSerSerSerProSerLeuVa 236  
QY 537 TTCTCGGGGGTGGCCA----- 552  
Db 236 lSerArg-----ProProArgLeuSerArgLeuAlaAlaLeuArgSerAspCysLe 254  
QY 553 -----CGTACACCGACAAGTTGTTTCAGTTCGCAATAACCGGTGGGAAGACATCCTG 605  
Db 254 uProLeuArgAlaProProArgLeuSerArgSerAlaLeuLeuThrSerSerThr 272  
QY 606 AGCGATGAGGTCAACGTGGCCGGTGGTGGCCAGCCTCT----- 645  
Db 273 -----ProTirpLeuIleAsnTirpProProSerAlaLeuSerSerAlaProVa 288  
QY 645 ----- 645  
Db 288 lAlaThrLeuArgProAlaLeuProGluSerArgProLeuProLeuLeuSerArgGluAl 308  
QY 646 -----TTGCCGGAGCTCTGTGTGGCCTGTGG----- 672  
Db 308 aLeuAlaSerValArgAlaProGluThrValProSerAlaTirpLeuSerSerAlaPr 328  
QY 673 ---ACGAAGAAGGCTCTGGACGCTACTCTATCTACATTGCCAATT----- 714  
Db 328 oSerThrArgSerMetProAlaArgLeuAlaArgValProLeuAlaLeuLeuSerAr 348  
QY 715 -----ACGCTTACGGTAATGTGGCCCTGATGGCCCTC 746  
Db 348 gArgAlaAlaAlaMetSerArgProAlaAlaProGluIleThrProAlaArgLeuLeuSe 368  
QY 747 ATTGAATGACCTGAGCCAGTG----- 771  
Db 368 rArgAlaTirpAlaMetArgValArgLeuAlaAlaProIleArgArgProProArgLe 388  
QY 771 ----- 771  
Db 388 uSerArgLeuAlaAlaAlaArgValIysAlaArgSerLeuGluIleSerProProAlaPr 408  
QY 772 -----ACCTCTCCGGGGCAATCTGGCGCTCAGAG----- 801  
Db 408 oLeuSerThrAlaProSer-----TirpArgSerSerSerAlaProGlyValAlaIleAr 426  
QY 801 ----- 801  
Db 426 gProProSerArgLeuThrSerValProProLeuArgSerArgValThrProValSerLe 446  
QY 802 -----ATNGGGCTG-----CTGAGCTGGGTGACCAATATACA 836  
Db 446 uThrArgThrProProArgTirpSerArgProPheArgLeuAlaSerSerArgProLe 466  
QY 837 GGGGGCCGAGCGGTACGCTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGC 896  
Db 466 uAlaLeuThrArgProCysTirpProLeuSerSerArgAlaAlaThrArgValMetProAl 486  
QY 897 GACAATGAGNATGGC-----CTAATTCTTTTCCACACCGGGCGATGGC 944  
Db 486 aLeuLeuProMetProProLeuArgLeuLeuSerSerThrProValArgIleSerThrAl 506  
QY 945 ACCTTTGTGGACCTGGCGGCTGTGTGTGGAGC-----AC 983  
Db 506 aProLeuAlaLeuIleThrProGluArgLeuLeuSerArgLeuAlaProCysSerValTh 526  
QY 984 CCCACCATGGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTATGGCAAGTGGAC 1043  
Db 526 rProAlaSerLeuSerSerLeuProProTirpLeuSerSer-----AlaAlaTirpLe 543

QY 1044 ATCGTCTATGTCACACTGGAATGGCCCCACCGCTCTATCTGCAATGAGCACCC---AT 1100  
Db 543 uValSer-----AlaSerAlaProAlaLeuGluIysValProProArgLe 558  
QY 1101 GGGAGGTCCGCTTCGGGGACATCGCTCACCAGTTCTCCATGCCCTCCCTGTCGCG 1160  
Db 558 uSerArgArgAlaAlaArgAlaValArgLeuProSerLeuThrSerAlaProProTirpLe 578  
QY 1161 ACGGTCTACACCGCGCTTTCGCAATGACACGAGCTGGAGATCTTCT----- 1209  
Db 578 uSerSerThrProProArgLeuThrLeuArgLeuPheTirpLeuSerSerArgProProSe 598  
QY 1210 -----TCAACA----- 1215  
Db 598 rProLeuAsnSerSerProProSerArgLeuArgProSerArgProAlaSerThrProLe 618  
QY 1216 -----ACATTGCC 1223  
Db 618 uAlaTirpPheSerArgArgCysThrValArgArgProLeuSerProMetThrLeuPr 638  
QY 1224 TACCGCAG----- 1231  
Db 638 o-ProArgLeuSerSerCysSerArgAlaPheThrAlaThrPheGluValLeuGluIles 658  
QY 1231 ----- 1231  
Db 658 erProAlaArgLeuSerThrCysArgAlaSerIleAlaMetProProPheAlaIles 678  
QY 1232 -----CTCTCA 1238  
Db 678 erGlnAspTirpLeuLeuSerIleValSerAlaAlaThrPheArgValPheSerLeuThrS 698  
QY 1239 GCCAACCGCCT-----CTTCGG-----CGTCATCCGTAGAGAGAC 1274  
Db 698 erSerProProCysTirpAlaArgLeuProSerValAlaCysArgSerProLeuAlaAla 718  
QY 1275 GGAGACCCCTCATCGA-----GGAGCTCAATCCGGCGAG----- 1311  
Db 718 leArgProProAlaLeuProThrAlaPheAlaAlaArgSerArgProSerLeuAsnS 738  
QY 1312 -----CCTTGAGCTGAGGCGGGGCACAGGGGGTGTGGTGCACGACTTCGACGG 1363  
Db 738 erLeuProProTirpSerLeuArgLeuAla-----ThrSerThrA 752  
QY 1364 AGACGGGATGC----- 1374  
Db 752 rgThrProAlaLeuProLeuAlaLeuProProSerArgLeuSerLysAlaProProVala 772  
QY 1375 -----TGA 1378  
Db 772 laLeuLysProAlaSerAlaThrSerThrProProArgLeuSerArgLeuAlaProTirp 792  
QY 1379 CC-----TCATCTTGTCCCATGG 1396  
Db 792 hrPheAsnSerLeuArgLeuAspSerArgProProAlaArgLeuSerSerCysProThrV 812  
QY 1397 AG-----AGTCCATGCTCAGCGCTGTCCGTCTTCCGGGGCAATCAGGCTTCAA 1447  
Db 812 alThrArgCysAlaProTirpLeuSerArgLeuProSerSer-----ArgLeuSerS 829  
QY 1448 CAACAACCTGGCTGCGAGTGGTCCACCGACCCGGTGTG-----GGGCTTTGCCCA- 1497  
Db 829 erArgProProArgLeuThrSerArgProProArgLeuLeuSerValProLeuProLeuL 849  
QY 1498 -----GGGAGCTAAGGTCTGCTCTACACCAAGAGAG 1531  
Db 849 euSerArgLeuArgAlaArgArgSerArgArgLeuAlaAlaGluAsnLysProArgArgL 869  
QY 1532 TGGGGCCCACTGAGGATCATCGAGGGGGCTCAGGTACCTGTGTGTGAGATGAGGCCGT 1591  
Db 869 eu-----SerMetAlaProLeuAlaThrThrSerArgSerSerPro- 882

```
QY 1592 GGCACACTTTGGCTGGGGAAGGATGAAGCCAGCAGTGTGGAGTGACGTGGCCAGATGG 1651
Db 882 -----
QY 1652 CAAGATGCTGAGCGGAAAGTGGCCAGCGGGAGAGTGAATCAGTGTGGAGATCCTCTA 1711
Db 883 -----
QY 1712 CCCCGGGATGAGGACACACTTCAGGACCCAGCCACTGGAGTGTGGCCAGGATTCCTC 1771
Db 885 -----
QY 885 euProArgLeuSerArgLeuProThrArgAlaSerThrArgLeuProGluThrSerP 905
QY 1772 CCAGCAGAAATGCCATTGCGATCGACACCAATGAATGCATCCAGTTCCTCCATTGCTGTG 1831
Db 905 roSerProArgLeuThrSerAlaTrp-----AlaLeuThrArgValAspCysT 921
QY 1832 CCTCGACAGACCCCGGTATGTCAACACCTATGGAAGCTACAGTGCCTGGGACCAACAA 1891
Db 921 rpLeuProSerArgPro-----CysTrpProLeuSerSerValAlaAlaProThrSerA 939
QY 1892 GAAGTGCAGTCGGGCTACGACCCCAACGAGATGGCA----- 1929
Db 939 rgSerPheArgProIleThrTrpProProArgLeuSerArgSerProArgValLeuThrT 959
QY 1930 -----CAGCTGCGTGGGCTGTGTGGAGCCCTGTGTGAAGAT 1966
Db 959 hrAlaArgSerArgLeuCysAsnArgProSerTrp----- 970
QY 1967 AGTGACACCAAGTTGGGAAGAGCCTTG----- 1995
Db 971 -----ArgLeuSerArgLeuLeuAlaSerThrArgSerSerProSerAlaSerS 987
QY 1996 --GTCCTGAATCACTGAATCACTGCTTGAATCAGCCGCT 2034
Db 987 erLeuProProArgLeuSerArgCys-----ProPro 997

RESULT 6
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30843
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Alignment Scores:
Pred. No.: 2,04e-08 Length: 663
Score: 217.00 Matches: 163
Percent Similarity: 29.69% Conservative: 57
Best Local Similarity: 22.00% Mismatches: 255
Query Match: 5.35% Indels: 266
DB: 4 Gaps: 37

US-09-914-958B-35 (1-2177) x US-09-252-991A-30843 (1-663)
QY 1 CGGAGAGCTGAGCCAGCCAGCCGAGCCGCGGTGGGAGCAGCAGCGCGCGCGCGCG 60
Db 17 ArgAlaAlaArgAlaSerPro-----TrpSerThrThrThrArgProProAla 32
QY 61 GCAGAGCGGCGAGCGAGCGCGCGCTTCCCA----- 90
```

```
Db 33 AspArgSerGlyAlaThrAlaArgValProAlaCysArgValProIleArgCysAla 52
QY 91 -----CGCCCTAGCGCGCGCGAGCGGGAGGATGGTCCGAGCG 135
Db 53 AsnAlaTrpProGlyArgProMetSerSerThrSerArgArgProGlyTrp----- 69
QY 136 CTGACCCCGCAITGTCAGAGATGTTACCGTTCTCTGCTGCTGTGGTTTCTGCCCATCA 195
Db 70 ---SerProAlaValPro-----AlaGlyAlaCys----- 78
QY 196 CTGAGGGTCCAGCGGCTGAACCCATGTTCTGCTGAGTCAACCACTCAGTTCCTGCTC 255
Db 79 -----CysTrp 80
QY 256 CTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGTGGCAGTTACTGATGTGGACC 315
Db 81 LysThrArgAsnValAlaGlyArgSerAlaThrAspAlaTrpSer----- 95
QY 316 ATGATGGGACTTTGAGATCGTGTGGCGGGGTACAATGGACCCCACTGGTTCGAACT 375
Db 95 ----- 95
QY 376 ATGACCGGCGCCAGAGCGGCTGGTGAACATCGCGTGCATGAGCGCAGCTCACCTACT 435
Db 96 -----SerAlaProAla 101
QY 436 ACGCGCTCGGGACCGGAGGAAAGCCATTGGGTGCACAGCTGCGGACATCGACGGG 495
Db 102 AsnCysCysCysLeu-----PheProAlaGlyProSerProAlaSerProAlaPro 118
QY 496 ACGCGCGGGAGGATCTACTTCTCAACACCAATAATGCTCTCTCGGGG----- 546
Db 119 AlaAlaCysArgArgTrpProArgAlaAlaCysHisTrpProAlaSerAlaTrpTrp 138
QY 547 -----TGGCCAGTACACCGCAAGTTGTTCAAGTTCGCGCAATA 585
Db 139 LeuAlaProAlaArgCysCysTrpProAlaProProAlaProAlaSerAlaGlyArgAla 158
QY 586 ACCGGTGGGAAGACATCTCGAGGATGAGTCAACGTGGCGCGCTG----- 630
Db 159 CysCysAlaSerProAsnArgArgArgGluProTrpProProSerProTrpAlaSer 178
QY 631 ---GTGTGGCCAGCTCTTTTG-----CCGAGCGCTCTGTGGCTGTGTGGAGCAGAAAG 681
Db 179 ArgAlaGlyProAlaSerCysGlyArgProProAlaCysSerProVal----- 194
QY 682 GCTCTGGAGCTACTCTATCTACATTGCCAATTACGCTACGTAATGTGGCGCTGATG 741
Db 195 -----AlaThrAlaProThrAlaThrCysSerPro----- 204
QY 742 CCTCATTTGAAATGGACCTGAGCGCAGTCACTCTCCCGGGGCAATCTGCGCTCAGAG 801
Db 205 ProSerAlaArgSerAlaTrpLysProCysAlaCysAlaLysAla----- 219
QY 802 ATGTGGCTGTGAGGCTGGGTCAGCAATAATATACAGGGGCGGAGCGTGTGGTGGGCC 861
Db 220 -----ValGlySerAlaArg-----SerProAlaAsnAlaTrpPro 231
QY 862 CCATCTGACGACAGTGTCTCGGATATCTTCGCGACATAGATGGGCTTAACCTCC 921
Db 232 AlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgPro-----SerAla 247
QY 922 TTTTCCACACCGCGGCGATGGCACCCTTTGTGGACGCTGGCGGCGAGTGTGTGTGGACG 981
Db 248 IleAlaSerThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThr 267
QY 982 ACC-----CCACACAGCATGGGCGAGTGTGCGCTCGCTGCTCACTCA 1023
Db 268 ThrThrProProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpProTrp----- 284
QY 1024 ACGGTGATGCAAGTGACATCGTCTATGGCACTGGAATGGCCCGCCCGCTCTATC 1083
```

```
Db 285 -----Ttp-----LysAlaProSerProAlaMetProProSerThrSerAla 298
QY 1084 TGCAAAATGAGCACCATTGGGAAGTCC-----GTTCCGGGACATCGCCT 1128
Db 299 ThrArgLeuAlaAlaCysGlyHisAspGlyAspAlaGlyArgAlaSerGlnThrProTrp 318
QY 1129 CACCAAGTTCTCCATGCCCTCCC-----CTGTCGCGCAGGTATCACCGCCACTTGG 1182
Db 319 ArgAspThrSerProCysValProAsnCysValAsnTrpArgArgArgilePro----- 336
QY 1183 ACAATGACAGGAGCTGGAGATCTTCTTCAACACATTCGCT----- 1224
Db 337 -----TtpSerAlaValAlaLysThrSerProTrpProArgTrpProGly 351
QY 1225 ---ACCGCAGCTCTTCAGCCACCGCTCTTCCCGTCATCTCCGTAGAGACGACGAGCC 1281
Db 352 MetProAlaGlyProArgProSerCysThrAlaValAlaAla-----Ttp 366
QY 1282 CCCTCATCGAGGAGCTCAATCCCGCGCAGCGCTTGGAGC----- 1320
Db 367 AlaProAlaArgGlyGlyTyrAlaAlaProArgAsnSerCysSerAlaGlyArgLeu 386
QY 1321 -----CTGAGGCGCGGCGCACAGGGGTGTGGTGACCG----- 1353
Db 387 ArgArgArgValArgArgSerAlaArgAlaTrpArgProTrpProAlaGlyArgAla 406
QY 1354 -----ACTTCAGGAGAGCGGATGCTGGACCTCATCTTGT 1389
Db 407 ThrProAlaSerAspAlaArgAspSerAlaSerArgGlyArgProAlaThrAlaAla 426
QY 1390 CCCATGGAGAGCTCATCGCTCAGC-----CGTGTCCGCTTCCGGGGCAATCAGGCT 1443
Db 427 AspHisProAlaAlaTrpValSerAlaAlaArgArgThrSerSerAlaProIleAlaGly 446
QY 1444 TCACCAACAACACTGGCTGCGAGTGTGCCACGCCCGGTTTGGGGCTTTGCCAGGGGAG 1503
Db 447 Ser-----AlaProGlySerGlyThrAlaPro----- 455
QY 1504 CTAAGTCTGCTCTACACAAGA-----GTGGGGCCCACTGAGGATCATCGACG 1557
Db 456 -----ArgCysHisProValArgLysAspGlyAlaGlyPro-----AlaSerThr 470
QY 1558 GGGGCTCAGGCTACCTGTGTGAGATGG-----AGCCGFGGCACACTTTGGGCTGGGA 1611
Db 471 GlyArgSerArg-----ArgTrpAlaAlaProAlaAlaArgAlaArgThrGly 488
QY 1612 AGATGAAGCCAGCAGTGTGGAGGTGACGTCGTCGTCGATGGCCAGATGCTGAGCCGGAACG 1671
Db 489 IleArgArgThrAlaAla----- 494
QY 1672 TGGCCAGCGGGAGATGAACACTAGTCTGGAGATCCTCTACCCCGGGATGAGGACA--- 1728
Db 495 -----AlaAlaCysTyrProProProAlaProGlyArgArgThrPro 508
QY 1729 -----CACTTTCAGACCCAGCCCACTGGAGTGTGGCCAGGATTCCTCC 1773
Db 509 AlaAlaArgArgSerAlaArgThrAlaProArgProAlaIleProAlaHisArgPro 528
QY 1774 ACAGAGAAATGGCCATTGCATGG-----ACACCAATGAATGCATCC 1815
Db 529 SerAlaSer---AlaThrGlyTrpProGlyProAlaLeuAlaCysProAlaAlaAlaGly 547
QY 1816 AGTTCCCATTCGTGTCCTTCGAGACAGCCCGTATGTGTCAACACTTGAAGACTACA 1875
Db 548 ArgArgProSerProAlaProAlaProAlaProAlaProHisAlaArgProAlaPro 567
QY 1876 GGTTCGCGGACCAACAAGATGCACTGCGGGCTTACGAGCCCAACGAGGTGCACAGCCT 1935
Db 568 ArgArgArgProGlyArgSerPro-----AlaThrAlaProAlaAlaPheAlaPro 585
QY 1936 GCG 1938
Db 586 Ala 586
```

```
RESULT 7
US-09-252-991A-21881
; Sequence 21881, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21881
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21881

Alignment Scores:
Pred. No.: 2,76e-08 Length: 720
Score: 215.50 Matches: 191
Percent Similarity: 30.83% Conservative: 39
Best Local Similarity: 25.60% Mismatches: 261
Query Match: 5.31% Indels: 255
DB: 4 Gaps: 41

US-09-914-958B-35 (1-2177) x US-09-252-991A-21881 (1-720)
QY 20 CCGGAGCCGGGCTGGGAGCAGCAGCGCGCGCGCGCGGAGCGCGGAGCGCGGAGCGG 79
Db 92 ProGlyProArgSerGluArgArgGlnArgArgAlaTyrArgArgGlnGlyArg 111
QY 80 CCGCTTCCACGCC-----CCTAGCGCGCGCGCGCGCGCGGAGCGCGGAGAT 124
Db 112 ProAlaGlyAspAlaTrpGlnProGlyProAlaArgTrpArgThrGlyGlnArgLysAsp 131
QY 125 GGTTCGAGGGTGACCCCGGATGTCAGGATGTTACCGTTCTCGTCTGCTGTGGTT 184
Db 132 ProGluGlyArg---GlyArgHis-----AspArgGlnGlnProValLeuAla--- 146
QY 185 TCTGCCCATCACTGAGGGGTCCCGAGCGGGTGAACCCATGTTCACTGCAGTCACCACTC 244
Db 147 -----AspArgGluTrpArgHisArgTyr 155
QY 245 AGTTCTGCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGCAGTTAC 304
Db 156 ArgGlyAlaAla----- 159
QY 305 TGATGTGGACCATGATGGGACCTTTGAGATGTCGTGGCGGGGTACAATGGACCAACCT 364
Db 160 -----GlyArgGlnProArgArgAspPheGlnHisLeuGluGly 173
QY 365 GGTTCCTGAAGTATGACCGGGCCAGAGCGGCTGCTGTAACATCGC-----GGTCAATGA 418
Db 174 ArgArgGlu-----ArgProArgAlaThrGlyGlnProArgArgGlnGlnHisArg 190
QY 419 GCGCAGCTCACCTACTACGCGCTCGGGACCGGAGGGGAAACCGCATTTGGGTACAGC 478
Db 191 ArgGlnArgProAlaProHisArgAlaProAla---GluProGlyGlnGlyGlyAla 209
QY 479 CTGCCACATCGACGGGAGCGCGGAGGAGATCTACTTCTTCAACCAACAATATGCTT 538
Db 210 GlyArgAlaArgArgProGluProGlyArgArgThrValGlnArg----- 226
QY 539 CTCGGGGTGGCCAGTACCGGACAGTTGTTCAAGTTCGCCAATAACCGTGGGAAGA 598
Db 227 -----ArgArgSerAlaArgGlnProGluGlnProGluArgGlu 239
```

QY 599 CATCTGAGCATGAGGTCAACGTGGCCGCTGTGTGGCCAGCCCTCTTTTGGCGGAGCTC 658  
Db 240 ProArgArgArgAlaGlyGlnProGlyArgArgAlaGlyGlnArg-ArgGlnPro 257  
QY 659 TGTGGCCCTGTGGACACAAAAGGGCTC---TGGAGCTTACTCTTACTATCTACATTGCCAATTA 715  
Db 258 AspGlyAlaArgArgAlaProGlyGlnProTrpArgAspLeu-----Leu 272  
QY 716 CGCTACCGTAAATGTGGCCCTGTGAGCCCTCAATTGAATGAGACCTGAGCCAGTGCCT 775  
Db 273 GlnArgArgCysAlaGlyAlaAspGlnPro-----GlyArgAlaGlyGln----- 287  
QY 776 CTCCCGGGCATCTGGCGCTCAGAGATGT----- 805  
Db 288 ---ProGlyArgGlnAlaAlaGlnArgCysArgArgAspAlaAlaGlyArgGlnProArg 306  
QY 806 -----GCTCTGAGGCTGG----- 820  
Db 307 GlnGlnProPheArgArgAspGlnArgGlnGlyCysGlyGlyTyProHisArgArgAla 326  
QY 821 GGTGAGCAAAATATACAGGGGGCCGAGCGCTGAGCTGGGGCCCATCTCT----- 868  
Db 327 GlyGlnGlnProGlnArgHisArgGlnGlnArgArgHisHisProGlyGlyArgPro 346  
QY 869 -----CAGCAGCAGTGCCTCGGATATCTTCTGGACAA 901  
Db 347 AlaGlyGlnArgProAlaGlyThrGlyGlnArgGln-----GlyProAlaArgGln 364  
QY 902 TGAGAAATGGCCCTAACTCTCTTCCACACCGGGCGGATGCACCTTTGTGAGCGTGC 961  
Db 365 ProGlu-----ArgProArgProAlaArgArg-----ArgPro 376  
QY 962 GCCAGTGTGTGTGGAGCAGCCCAACAGCATGG-----CCGAGGTCT 1006  
Db 377 GlyGlnArgAsnArgArgHisProArgProGlnTrpArgHisAlaGlyGlnProArg 396  
QY 1007 CGCCCT-----GGCTGACTTCAACCTGTGTCGCAAGTGCACATCGT 1048  
Db 397 ArgProAspArgHisAlaArgArgAlaAlaAlaProAlaArgArgGlyGlyGlnArg 416  
QY 1049 CTATGCAACTGGAATGGCCCGCCCGCTCTATCTGCAATGAGCACCA----- 1099  
Db 417 ArgArgArgGlyAsnLeuGlnArgProArgLeu-----HisProArgArgArg 432  
QY 1100 -----TGGGAAGTCCGCTTCGGGACATCCG 1126  
Db 433 GlnProGlyGlnProArgArgAlaProAspArgArgGlnProAspProAlaHisArg 452  
QY 1127 CTCACCAAGTCTCCATGCCCTCCCTGTCCGACGGTTCATCACCAGCTTTTGACAA 1186  
Db 453 -----ProGlyProGly-----Gln 457  
QY 1187 TGACCAAGAGCTGGAGATCTTCTCAACAACATTCCTACCGCAGCTCTCAGCCAAACCG 1246  
Db 458 GlnProGlyArgGlyAspLeuArgArgArg-----ProGlyHisArgGlyArgSer 475  
QY 1247 CCT----- 1249  
Db 476 ProGlyGlnGlnArgGlnGlyHisProGlyGlnProArgHisArgProAlaArgArg 495  
QY 1250 -----CTTCGGCTATCCGTAGAGACGACGAGCCCGCTCTATCAGGAGCT 1297  
Db 496 ArgArgAlaGlyGlnProArgArgArgHisArgPheArgArgPro----- 511  
QY 1298 CAATCCCGGCGAGCTTGGAGCTGAGGCGCGGGGACAGAGGGGTGTGTGACCGACTT 1357  
Db 512 AspAlaArgGlnArgLeu-----ProGlyGlnGlnArgGlnGlyProAlaLeu 527  
QY 1358 CGACGAGAGCGGATGTGGACCTCTCTTGTCCCATGGAGTCT-----CATGGCTCA 1411  
Db 528 ArgGlnArgArgProGluArgArgHis-----TrpArgAlaGlyGlnArgArgGly 544  
QY 1412 GCGGCTGTCGCTTTCGCGGGGCATCAGGCGCTTCAACAACTGGGTGCGAGTGTGC 1471

Db 545 TrpProValAspGlnProGlyArgAlaGlyArgGlnGlnArgProArgGlnProTrp 564  
QY 1472 ACGCACCCGCTTTGGGCGCTTGGCCAGGGAGCTAAGGTGCTGTCTACACCAAGAGAG 1531  
Db 565 ArgArgProGlnTrpGlnAlaValAlaPro---GluArgArgGlnProGlyGlnPro 583  
QY 1532 TGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGT 1591  
Db 584 TrpArgProAlaAspGlnArgArg-----Thr 593  
QY 1592 GGCACACTTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGG 1651  
Db 594 GlyThr----- 595  
QY 1652 CAAGTGTGTAGCCGGAACGT---GGCCAGCGGGGAGATGAACCTCAGTGTGTGAGATCCT 1708  
Db 596 ---AspGlyArgAlaArgLeuArgArgArgArgAsnLeuArgProGlyArgPro 614  
QY 1709 CTACCCCGGGA-----TGAGGACACACTTTCAGGACCCAGCCCTGAGTGTGGCCA 1762  
Db 615 AlaProAspGlyArgAlaProGlyAlaThrProGlyProAlaGlyArgArgAlaArg 634  
QY 1763 AGGATTTCCACAGCAGGAAA-----TGCCCATTCGATGACACCA 1804  
Db 635 GlnSerArgProAlaGlyArgArgProGlyGlnProGlyTrpPro-----AspGln 651  
QY 1805 TGAATGCATCCAGTTCCCTTCCCTGCTGCTCCAGACAGCCGCTATGTCAACACCTA 1864  
Db 652 CysSerArgProAlaGluHisArgAlaAlaGluArgProArg----- 665  
QY 1865 TGGAAAGTACAGTCCGCGACCAACAAGAAAGTGCAG---TCGGGGCTACGAGCCCAACGA 1921  
Db 666 -----GlnProGlyArgArgAspPheGlnProAlaGlyLeuArgAlaAlaGly 682  
QY 1922 GGATGCGCAGCTCGGT 1939  
Db 683 GlnAlaHisArgGlnArg 688  
RESULT 8  
US-09-252-991A-18995  
; Sequence 18995, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18995  
; LENGTH: 595  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18995  
Alignment Scores:  
Pred. No.: 2,77e-08 Length: 595  
Score: 215.00 Matches: 165  
Percent Similarity: 32.82% Conservative: 48  
Best Local Similarity: 25.42% Mismatches: 223  
Query Match: 5.20% Indels: 214  
DB: 4 Gaps: 37  
US-09-914-958B-35 (1-2177) x US-09-252-991A-18995 (1-595)  
QY 1725 CCTCATCCCGGGGTAGAGGATCT---CCAGCACTGAGTTCACTCCCGCTGGCCACGT 1669

Db 38 ProGlyProArgArgGlyAsnArgProArgLeuArgThrSerArgArgThrArgGly 57  
QY 1668 TCGGCG-----TCACCATTTGCCATCTGGCCACGTCACCTCCA 1630  
Db 58 ArgGlyArgGluArgThrProAlaGluSerProAlaCysAlaLeuArgThrValProArg 77  
QY 1629 CACTGTCTGCTTATCTCTTCCCGAGGCAAAAGTGTGCCAGGGCTCCATCTCACACAGT 1570  
Db 78 CysArgTyr----- 80  
QY 1569 AGCTGAGCCCGCTGCATGATCTCAGTGGGCCCCCACTCTTCTGCTGAGCACGA 1510  
Db 81 -----AlaGlyAspArgArgArg 86  
QY 1509 CCTTAGTCTCCCTGGCAAGGCCCA-----ACC 1480  
Db 87 ProValThrValPheProArgProGluLeuValHisArgArgArgThrAla 106  
QY 1479 GCGTGGCTGGCACCACTCGCAGCAGTGTGTGTGAAGCCCTATTGCCCGGAGACGG 1420  
Db 107 GlyAlaArgPro-----ArgGlyThrArg 115  
QY 1419 ACAGCGCTGAGCATGAGTCTCCATGGGCAAGATGAGTCCAGCATCCCTCTCCGT 1360  
Db 116 ProAlaGlyAlaAlaLeuSerGlyThrGlyArgProGlyAlaThrGlyArgValGln 135  
QY 1359 CGAAGTCGGTCACACACCCCTGTGC-----CCGGGCTCAGGCTCCAAAGGCTCCGCGG 1303  
Db 136 ArgAlaArgValProGlnArgGlnAlaArgProGlyArgSerHisArgArgSer 155  
QY 1302 GATTGAGCTCTCGATGAGGGGTCTCCGTGCTCTACGGATGACCGGCAAGAGCGGT 1243  
Db 156 Asp-----ArgGlyGlnLeuGly 161  
QY 1242 TGCGTGAAGGAGC-----TGCGGTAG-----CAATGT 1216  
Db 162 AlaGlyArgAlaGlnCysIleThrPheProAlaGlyArgIleLeuThrProGlyThrCys 181  
QY 1215 TGTGTAAGAGATCTCCAGTCTCTGTGAT-----TGTCAAAGTCGGCGGTGATACCGTGC 1159  
Db 182 IleAspArgAlaThrAspPheProAlaHisLeuCysArgSerArgHisArgLeuProArg 201  
QY 1158 GGACAGGGAGGCGATGGAGACTTGGGTGAGCGGATCTCCCGA-----AGCGACCTTCC 1102  
Db 202 GlyArgAsnArgLeuProArgArgTrp-----ProCysThrGlyAlaSerGlyLysSer 219  
QY 1101 -----CATGGGTGC-----TCATTTCAGATAGA-----GGCGGTGGGGC 1066  
Db 220 ThrHisArgIleIleHisSerAlaAlaArgSerPheProArgArgProAlaAlaArgArg 239  
QY 1065 CATTCCAGTTGCCATGACGATGTCCACTTTGCCATCAGCGTTGAAGTCAGCCAGGCGGA 1006  
Db 240 HisAspArgGlyHisArgArgThrThrGluCysArgGln-----ValGlnProAlaGlu 257  
QY 1005 CACCTCGCCCATGCTGGTGGGGTCTCCACAC-----CAGCACTGGCGCAG----- 958  
Db 258 CysProGlyArgSerArgGlyHisArgHisGlyHisSerArgHisHisProArgCys 277  
QY 957 -----CGTCCAAAGGTGCCATCGCCCGGTGTGGA----- 925  
Db 278 ThrAlaArgThrTyrProHisArgTrpHisAlaProAlaCysGlyHisArgArgPro 297  
QY 924 -----AAAGGAAGTTAGGCCCATCTCTATTGT 898  
Db 298 AlaGlyTyrArgGlyProCysGlyLysAspArgArgGlyThrCysAlaGluSerHis--- 316  
QY 897 CGCAGAAGATATCCAGGACACTCTGCTGAGGATGGGCGCCAGCTGAGCCCTCGCCGCC 838  
Db 317 ---ArgArg-SerArgSerArgAlaAlaGlyGlyArgHisArgProGlySerSerG1 335  
QY 837 CTGTATATTGTCTGA-----CC 821  
Db 335 yProValPheProMetAlaGlyValSerArgProAlaThrGlyThrGlyGlnGlyHisPr 355

QY 820 CCAGCCTCAGCAG-----CCACATCTCTGAGCGCCAGAAATGCCCGGAGAGGTCA 770  
Db 355 oAspProGlnGlnGlyPheValHis-----ArgValHi 367  
QY 769 CTGGCCTCAGGCT-----CCATTCAATGAGGGCATCAGGGCCCACTATCCG 722  
Db 367 sTrpProArgGlyLysArgGlyProArgHisHisProLeuGlyAlaHisArgAr 387  
QY 721 TAGCGTAATTGGCAATGTAGATAGTCCAGAGCCCTTCTGTCTCCACACAGGCC 662  
Db 387 sArgProGly-----AlaAlaAlaArgThrProGluGlyLeuHisGlyLe 402  
QY 661 ACAGAGCTCCGGCAAGAGGCTGGCCACACACCGGGCCACGTTGACTCATCGCTCAGG 602  
Db 402 uArgThrAspArgGlyLysArgPhe-----GlnArgProAlaProLeuG1 418  
QY 601 ATGTCTTCCACCGTTATTGCGAACTTGAACAACCTGTCCGGTGTACGTGGCCACCC 542  
Db 418 ySerAlaThrProGlyArgAlaGlyAlaGlyAlaTrpProGlnProThrTyrProGlnTr 438  
QY 541 GAGAAGCGCATATTGTGTGAGGAAGTAGATCTCTCCCGCGCGTCCC-----CGTCGATG 485  
Db 438 pArgArgArgThrAlaGlyArgGlySerAlaProGlyProAlaAlaProGlyArgAsnHi 458  
QY 484 TCCAGGCTGTGACCCCAATGGCTTCCCTCCCGCGTCCCGCAGCG----- 439  
Db 458 sArgArgIleHisPro-----ArgArgProAlaGlySerHisLeuLeuGluPheLeuHi 476  
QY 438 -----CGTAGTAGGTGAGCTG 422  
Db 476 sArgGlnValMetGlyLeuGlyProArgArgValValAspAsnGlnGlyLysValArgCy 496  
QY 421 CGCTCATCGACCGCATGTTCCACAGCGCTTCTGGCGCGGTCTATCTTCAGAACCCAGG 362  
Db 496 sValLeuSerPheHisCysAlaProCysArgArg-----HisSerThrThrGlnAl 513  
QY 361 TTGGTCTCATGTTACCCCGCCAGCAGCATCTCAAAGTCCCATCATGTCATCATCAGTA 302  
Db 513 aTrpIleAlaGlyAlaProLeuArg-----ProArgGluAsn-HisLeuArgA 529  
QY 301 ACTGCCACACCATAGTTGAGCTGGGTGGGATTAATCTCATAGTCAGGAGCAGA----- 248  
Db 529 rGysArg-Ser-----AlaTrpAspGlyTyr-----ProProGlyArgGlyLeu 543  
QY 247 -----ACTGAGTGTGCTGCTGACGTGCAATCGGTTTCAGCCCGCTGGGACCCTCA 197  
Db 544 CysThrGlyThrAlaLeuArgSerGlyValArgGluGlyAlaAlaLeuTrpSerIleSer 563  
QY 196 GTGATGGCGAGAAACAGAGCAGCAGCAGGAAACGGTAAACATCTCGACATGCGCGGTCA 137  
Db 564 IleCys-----SerProLeuSerAsnGly-----ValSerIleProLeuSer 577  
QY 136 GCGCTCGGAGCATCTCTCCCG 116  
Db 578 ArgLeu-----AlaLeuSerPro 583

## RESULT 9

US-09-252-991A-19127  
; Sequence 19127, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: ABRUQUINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 19127  
; LENGTH: 590  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19127

## Alignment Scores:

Pred. No.: 4, 3e-08 Length: 590  
Score: 212.50 Matches: 163  
Percent Similarity: 30.68% Conservative: 26  
Best Local Similarity: 26.46% Mismatches: 216  
Query Match: 5.14% Indels: 211  
DB: 4 Gaps: 36

US-09-914-958B-35 (1-2177) x US-09-252-991A-19127 (1-590)

QY 1518 AGAGCAGCAGCTTAGCTCCCTGGCAAGGCCCAACCGGGTGGTGGCCACCTCGCA 1459  
DB 7 ArgAlaArgPro-----AspProAlaGlyGlyAlaAlaGlyArgArg 21  
QY 1458 GCCAGTTGTTGAAGCCCTGATTGCCCGGAGCAGCAGCGGCTCAGCCATGGACT 1399  
DB 22 AlaAsp-----ProProAlaGlyAlaGlyArgAlaAlaAlaThrLeuProAspPro 38  
QY 1398 CTCCTATGGCAAGATGAGGTCCAGCATCCCGTCTCCGTCGGAAGTCGGTCACACACCCC 1339  
DB 39 -----GlyAlaGluProAlaProGlyThrArgGlnArgArgGlyArgSerAlaGlnSer 56  
QY 1338 CTGTGC-----CCCGGC-----CCTCAGGCTCCAAAGCGTCGCGGATGAGCTCTCGA 1288  
DB 57 GlnAlaArgProGlyArgProAlaAlaProGlyArgThrArgAla-----AlaArgArg 75  
QY 1287 TGAGGGGTCTCCGTGCTCTACGATGACCGGAGAGCGGTTGCTGAGGAGCTGC 1228  
DB 76 ProArgArgLeuAlaProGlyGlyAlaHisArgGlyAlaThrArgAspGlnPro 95  
QY 1227 GGTAGGCAATGTTGTAAGAGATCTCCAGCTCTCGTCAATGTCAAAGTCGGCGTGA 1168  
DB 96 GlyAlaGln-----ProGluArgGlyAlaProGlyHisGly----- 107  
QY 1167 TGACGTCGGACAGGGAGGCGATCGAGAACTTGGTGAGCGGATGTCGCGAAGCGGA 1108  
DB 108 ---ProGlyGlyGln-----ArgThrGlyVal-----ValGlyLeuGly 119  
QY 1107 CTTTCCCAT-----GGGTGCTCATTGTCAGATAGGCGGT----- 1072  
DB 120 ProGlyGluArgProGlyProAlaPheLeuSerArgGlyLeuGlnArgArgArgGly 139  
QY 1071 -----GGGGCCATTCCAGTTGCCATAGA-----CGATGTCACCTTTGCCAT 1030  
DB 140 ArgSerArgGlyLeuSerGlnAlaAspArgLysHisProProArgArgProAlaAlaHis 159  
QY 1029 CACGGT----- 1024  
DB 160 ProSerArgProArgArgThrSerGlnGlyAlaGlnArgAlaValProGlyValPro 179  
QY 1023 TGAAGTCACCGAGCGCACCTCCCATGCTGCTGGTGGGGTCGCTCA---CACACGAC 967  
DB 180 ArgAlaArgProGlyArgLeuAlaLeuAlaGlyGlySerTrpProGlyHisGlySer 199  
QY 966 TGGCCG-----CAGCGTCCACAAAGGTGCCAT 940  
DB 200 ArgProAlaGlyThrGlyAspAlaAspAlaArgHisAlaGlnArgHisGlnArgAlaGln 219  
QY 939 CGCCCGCGT-----TGTGAAAAGGAGTTAGGCCCATCTCTATTGTCGCAGA 892  
DB 220 AlaProGlyArgAlaAlaAlaThrGlyArgAspSerValArgGlyGlyGlnArgGluHis 239  
QY 891 AGATATCCGAGCAGCTGCTGCTGAGGATGGGCGCCACGCTGACGCTCGGCCCTCTGTAT 832  
DB 240 ArgAspProArgSerGlyLeuGlnProAlaGlyGlyGln----- 252  
QY 831 ATTTGTGACCCCGCCTCAGCAGCCACATCTCTGAGCGCCAGAAATGCCCGGAGAG--- 775

DB 253 -----ProGlyLeuGlnSerHisArgLeuSerProArgGlyArgGlyArgAla 269  
QY 774 -----GGTCATGGCTCAGGTCCTCATTTCAA----- 748  
DB 270 GlnArgGlnProAlaGlyGluTrpProArgGlyProAlaProLeuProAlaAspProArg 289  
QY 747 -----TGAGGCATCAGGCGCCACATACCGTAGGGCTAAT 712  
DB 290 ArgAlaAlaProAlaArgAlaValAlaGlyArgThrGlyGlyTyr-----AlaGluGlu 307  
QY 711 TGGCAATGTAGATAGAGTAGCGTCAGAGCCCTTTCTGCCACACAGGCCACAGAGCGTC 652  
DB 308 TrpArgSer-----LeuSerProMetAlaAlaAlaGlnProAlaAlaArg 322  
QY 651 CGGCA-----AGAGGTGGCCACACCGGCCACGTCACCTCATCTCGTCAGGATGCTT 595  
DB 323 ArgGlnTrpArgGlyAspProHisArg-----ArgPheLeuHisArgSerGlnCys--- 339  
QY 594 CCCACCGGTTATTGCGGAACCTTGAACAACTTGTCTCGGTGTACGTGCCACCCCGAGAGG 535  
DB 340 -----ProProArgArg 343  
QY 534 CATTATTGGTGTGAGGAAGTAGATCTCTCCCGCGCTCCCGTCGATGTCGAGGCTG 475  
DB 344 -----ArgGlyAlaProAlaSerProAlaAlaLeuArgArgThr 356  
QY 474 TGACCCCAATGGCGTTCCCTCCCGTCGCGCAGCGCTAGTAGGCTGAGTGGCGTCA 415  
DB 357 AspArgProGlyGlnProProValValProArgAlaAlaAlaArgGlyGlyProAlaGly 376  
QY 414 CGACCG-----CGATGTTCCACGACCGCTTCTGGGCCCGGTCACTTCA----- 370  
DB 377 AlaProGlyGlyArgAlaGlyProAlaAlaAspArgProGlyProLeuGlnAlaAla 396  
QY 369 -----GAACACAGTTGGTCCATTGTATCCCGCCAGCAGATCTCAAAGTCCCATCAT 316  
DB 397 GlnArgGlnProArgProArgGlyArgArgProValAlaGlnGlyAspGlyProAlaPro 416  
QY 315 GGTCCACATCACTAAGTCCACACCATAGTTGAGCTGG----- 277  
DB 417 AlaAlaHisArgAlaGlySerGlnHisProGlyAlaSerGlyArgArgValArgHis 436  
QY 276 -----TGGCATTAATGTCATAGTCAGGAGGAGAACTGAGTTGGTACTGCAGTGA 226  
DB 437 SerProArgTrpArgHisArgHisArgGlyPro----- 447  
QY 225 ACATGGTTACCGCTGGGACCCCTCAGTCATGGGAGAAACCCAGCAGCAGCAGGA 166  
DB 448 -----ValAlaProGlyArgThrPro-----AlaGly 456  
QY 165 ACGGTAAACATCTCGGACATGCCGGGTGACGCTCGAGCCATCTCTCCGCTCTCGGCC 106  
DB 457 ProValAlaPro-----AlaGlyGluArgAlaArgPro-----ArgValAspPro 471  
QY 105 CGCCCGCTAGGG-----CGTGGGAGCGGGC----- 79  
DB 472 ArgArgLeuAlaGlyAspGlnProValPheArgAlaGlyThrGlyAsnLeuArgAlaAsp 491  
QY 78 -----GCTCGCTCCGCTCTCGCGCGCGCGCGCGCGCGCGCTGCTT 40  
DB 492 GluProGlyGluProGlyAspAlaThrCysGlnAlaThrArgArgGlnTrpLeuProVal 511  
QY 39 GCTCCC-----AGCCCGCTCCCGGG 19  
DB 512 LeuProArgGlnProAlaGlyGlnHisProArgThrProAlaProGly 527

## RESULT 10

US-09-252-991A-27068  
; Sequence 27068, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;/ FILE REFERENCE: 107196.136  
;/ CURRENT APPLICATION NUMBER: US/09/252,991A  
;/ CURRENT FILING DATE: 1999-02-18  
;/ PRIOR APPLICATION NUMBER: US 60/074,788  
;/ PRIOR FILING DATE: 1998-02-18  
;/ PRIOR APPLICATION NUMBER: US 60/094,190  
;/ PRIOR FILING DATE: 1998-07-27  
;/ NUMBER OF SEQ ID NOS: 33142  
;/ SEQ ID NO 27068  
;/ LENGTH: 638  
;/ TYPE: PR1  
;/ ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27068

## Alignment Scores:

Pred. No.: 4,45e-08 Length: 638  
Score: 212.50 Matches: 188  
Percent Similarity: 29.81% Conservative: 26  
Best Local Similarity: 26.18% Mismatches: 263  
Query Match: 5.14% Indels: 241  
DB: 42 Gaps: 42

US-09-914-958b-35 (1-2177) x US-09-252-991A-27068 (1-638)

QY 1830 ACACGAATGGGAAGTGGATGATTCATTTGGTGTCCATGGCCATTTTCCTGCTGGG 1771  
DB 35 ThrArgHisGlyAlaLeuProValAlaArgProValProGlyAlaValAlaLeu 54  
QY 1770 AGAATCTTGGC-----CACACTCAGTGGGCTGGTCTGAAGTGTCTCAT 1720  
DB 55 ArgArgThrGlyArgAlaGlnHisProProVal-----ProArg 67  
QY 1719 CCGGGGGTAGAGGATCTCCAGCA---CTGAGTTTCATCCCGCTGCCAGCTTCGGC 1663  
DB 68 GlnArgArgArgAlaProAlaGlyArgAlaAlaSerProGlyLysProArgArgGly 87  
QY 1662 TCACCATCTTGGCATCTGCCAGCTCAGCTCCACACTGCTGGTTCATCTTCCCGCAGGC 1603  
DB 88 LeuAlaGly-----AlaThrAlaGlyPro-----ArgProAspPro--- 99  
QY 1602 CAAAGTGTCCAGGCTCATCTACACAGGTAGCTGAGCCCGCTGATGATCCTCA 1543  
DB 100 ArgAlaLysProArgGlyValAlaArgGlnGlnArgArgProAlaArgArgGln 119  
QY 1542 -----GGTGGGCCCACTCTCTTGTGTGTAGACACGACCTTAGCTCCCC 1498  
DB 120 AlaArgArgProAlaGlyProAlaAlaThrGly---ArgAlaArgPro----- 135  
QY 1497 TGGCAAGGCCCAACACCGGGTGCCTGGCCACCACCTCGCAGCCAGTTGTTTGAAGCCCT 1438  
DB 136 -----HisArgGlnProGlyThrMetValArgLeu----- 145  
QY 1437 GATTGCCCGGAGACGAGCAGCGGCTGACGCTAGGACTCTCATGGGCAAGATAGGT 1378  
DB 146 -----ProGlyArgArgGlyAlaAlaPro-----GlyArgArg---Leu 158  
QY 1377 CCAGCATCCGCTCCGTCGAGTGGTTCACACACCCCTGTGCCCCGCGCTCAG--- 1321  
DB 159 ProHisProAlaAlaArgGlnProAlaProAspProGlnProGlyProArgGln 178  
QY 1320 -----GCTCAAGCGCTGCGCGGGAATTGAGCTCTCGATGAGGG 1282  
DB 179 AlaGlyLysArgGlnProGlyProGlyArgArg----- 191  
QY 1281 GGTCTCGTCTCTCTACGATGACGGGAGAGCGGT-----TGGCTGAGG 1234  
DB 192 -----ArgAlaArgGlnGlyGlnProAlaArgAlaGlyArgAlaAspArgGlyLeuAla 209  
QY 1233 AGCTGCGGTAGGCAATGTTGTTGAAGAGATCTCCAGCTCTCGTCTTAAAGTCCG 1174  
DB 210 AlaAlaArgArg-----AlaArgArgAlaThrLeuGlyGlnArgArgLeuArg 225

QY 1173 CGGTGATGACCTGGCGGACAGGGGAGGGCATGGAGAACTTGGTGAGGCATGTCCTCGGA 1114  
DB 226 Arg-----HisAspGlyAlaGlyArgArgArgArgGlyTipArgArgGlnCys----- 241  
QY 1113 AGCGACCTTCCCATGGGTGCTCAATT----- 1087  
DB 242 ArgArgProArgArgGlyAlaAlaAlaArgThrArgGlnProAlaAlaThrLeuSerGly 261  
QY 1086 -----GCATAGAGCGGCTGGGGGCCATTCCTCAGTTCCTCATAGA 1048  
DB 262 ArgValAlaProHisProArgProAspArgAlaSerProAlaThrValArgArgHisArg 281  
QY 1047 CGA-----TGTCCACTTTCCTCATCCGTTGAAGTCAGCCAGG----- 1009  
DB 282 ArgGlnThrGlyArgGlyAlaPro-----GlyProGlyHisProGlyThrAla 297  
QY 1008 -----CGACACTCGCCCATGCTGTGGGGTCTGCCACACGACTGGCCGCGAG 958  
DB 298 GlyAlaArgArgThrProAlaProAlaArgProGlyThrProAspProGlyTrpAspAsp 317  
QY 957 CGTCCACAAAGGTGCTCATCGCCCGGTTGTGGAAGAAAGTATTAGGCCCATTTCTCATGT 898  
DB 318 ArgProAspProAlaHisArgPro----- 325  
QY 897 CGCAGAAGATATCCGAGGCACCTGCTCTGAGGATGGGCGCCACGCTGACGC----- 847  
DB 326 ArgHisArgGlnProAlaLysThrPro---GlyAlaGlyProArgAlaThrGlyAlaGly 344  
QY 846 -----CTGGCCCCCTGTATTTTGTGTGACCCCGCAG---CCTCAGACGACCATCTC 799  
DB 345 AlaLeuGlyLeuGlyArgArgArgLeuLeuAlaProHisIleProAlaTyrProHisArg 364  
QY 798 TGAGCGCCGAATGCCCGGGAGAGGTCACTGGCCT-----CAGGTCCTCATTTCAATGA 745  
DB 365 ---GlyProAlaGlnProGlyArgPheAlaGluProProAlaGlnLeuProArgArgAla 383  
QY 744 GGGCATCAGGGGCCACATTACCGTAGCGTAATTGGCAATGTAGATAGAGTACGCTCCAG 695  
DB 384 GlyArgArgAspPro----- 388  
QY 684 AGCCCTTTCTGTCACACAGGCCACAGAGCGTCCGCAAGAGGCTGGCCACACCGCGG 625  
DB 389 -----ProProArgArgThrGlyArgArgGlnGlnProAspProGlyArgGly 404  
QY 624 CCACGTTGACCTCATCGCTCAGGATGCTTCCACCGGTTATTGCGGAACCTTCAACAAC 565  
DB 405 GlnGlnArgProAlaArgArgGlyArgAlaProGly-----ArgArg 418  
QY 564 TGTGCGTGTACGTGGCCACCCCGAGAGGCATTTATGTTGTAGAGAAAGTAGATCTCCT 505  
DB 419 TyrArgAlaAspProArgArgProArgArgHis-----GlySerGlyAspPro 434  
QY 504 CCGGCGCTCCCGTCTGA-----TGTGCGAGCTGTGACCCCAA----- 466  
DB 435 AlaGlyArgArgArgGlnProAspArgArgGlnSerGlnArgArgProGlySer 454  
QY 465 -----TGG-----CGTTCCTTCCTGCGGTCGCGCGCTAGTAGGCGTG 427  
DB 455 GlyAlaThrProAlaGlyAspArgArgGluProGlyArgProAlaAlaArgAlaGlyArgG 474  
QY 426 AGCTGCGCTCATCGACCGCATGTTCCACGCGGCTTCTGGGCGCGCT----- 379  
DB 475 GlyAlaArgGlnArgProValAspArgAlaProGlyArgGlyValGlyAspHisArgGln 494  
QY 378 -----CATACTTCAGAACCGAGTGGTTCCTATCTACC 346  
DB 495 ArgAlaAspGlyAspProLeuAspArgArgThrAspGlnProAlaArgProGlnCysArg 514  
QY 345 -----CCGCGACGACGATCTCAA-----AGT 325  
DB 515 HisArgGlyArgProArgArgProGlyProArgLeuArgArgGlyArgArgGlySer 534





```
Db 416 euArgArgLeuArgLeuPro-----422
QY 690 GTCAGAGCCCTTTCTCTCCACACAGGCCACAGAGCTCCGGCAAAGAGGCTGCCACAC 631
Db 423 -----HisArgArgProValp 428
QY 630 CACGGCCACG-----TTGACCTCATCGCTCAGATGCTCT-----TCCAC--C 589
Db 428 roArgAlaGluProGlyLeuGlyHisSerArgIleSerAsnAspAlaSerHisArg 448
QY 588 GGTATTTCGGAACTTCAACAACCTGTCGGTGTACGTGGCCACCCCGAGAGAGCAATTAT 529
Db 448 rgSerLeuArgHisProGlyArgArg-ArgGlnArgIleProAlaArgArgProHisHis 467
QY 528 TGGTGTGAGAAAGTAGATCTCTCT---CCGGCCCGTCCCGTCGATGTCGAGGCTGTGA 472
Db 468 AspHisAlaAlaSerArgProArgProGlyThrGlyArgArg-----482
QY 471 CCCCAATGGCGTTCCCTCCCGGTCGCCGACGCGGTAGTAGGCTGAGCTGGCTCATCGA 412
Db 483 ProProArgArgArgSerLeuHisAlaAlaAspGlyGlnProGlyAlaAlaGluArg 502
QY 411 CCGCGATGTTTCACACAGCCGCTTCTGGGCCCGGTCATCTTCAGAACACAGGTTGGTCCAT 352
Db 503 ProArgArgArgProGlnProAlaGlyAlaGlyArgArgProThrAlaGlyAspAlaHis 522
QY 351 TGTACCCGCCACAGACATCTCAAGTCCCATCATGTCACATCATAGTAACAGTCCACAC 292
Db 523 ---ProPro-----AlaGlnHisArgPro-----530
QY 291 CATAGTTGACGTGGTGGATTACTGTCTAGTCAGGAGGAGCACTGAGTGGTGACTG 232
Db 531 ---SerGlyAlaGly-----534
QY 231 CAGTGAACATGGTTACGCCGCT-----GlnProAlaArgAlaThrAlaArgLeuProAlaProValArgGly 549
Db 535 -----GlnProAlaArgAlaThrAlaArgLeuProAlaProValArgGly 549
QY 207 ---GGAGCCCTCAGTGATGGCGAGAACCCAGACAGCAGCAGGAAACCGTAAAC-ATCCTG 152
Db 550 GluGlyThrProGlyArgArgGlyAlaGlyArgSerAspAlaGlyAlaLeuArgAlaLeu 569
QY 151 GACATGCCGGGTACGG-----CTCGAGACC 125
Db 570 HisArgProArgSerGlyAlaGlnHisAlaValGlyGlnHisGlnArgLeuGlyGln 589
QY 124 ATCCTCCCGCTCTCGGCCCGCGCTAGGGGGCTGGGAAAGCGGCGCTCGCTCGCGCG-- 68
Db 590 AlaGluProAlaAspHisProAlaGlnArgGlyLeuGlyArgArgGluGlyLeuProVal 609
QY 67 -----CCTCTGCCGCGCGCGCGCGCTGCT-----41
Db 610 AlaArgThrLeuProAlaGluProAlaProAlaProAlaProAlaGlyAlaAlaValPro 629
QY 40 TGCTCCAGCCCGGCTCCCGGCTGGGCTC 11
Db 630 ValHisGlnProArgLeuArgArgProLeu 639
```

## RESULT 12

```
US-09-252-991A-23774
; Sequence 23774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
```

```
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23774
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23774

Alignment Scores:
Pred. No.: 6,97e-08 Length: 798
Score: 210.50 Matches: 177
Percent Similarity: 30.42% Conservative: 49
Best Local Similarity: 23.82% Mismatches: 239
Query Match: 5.19% Indels: 278
DB: 4 Gaps: 40

US-09-914-958B-35 (1-2177) x US-09-252-991A-23774 (1-798)

QY 52 GCGCGCGGCACAGCGCGGACGAGCGCGGCTTCCACGCCCTAGGCGCGGGCGG 111
Db 3 AlaAlaAlaMetSerAlaGlyAlaSerAlaArg-----ArgCysTrpPro 17
QY 112 AGAGCGGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCAGGATGTTACCGTTCCTGC 171
Db 18 MetAlaGlySerMetAlaCysValAlaArgTrpThrPro---CysArgArgGlySer 36
QY 172 TGTGCTCTGGTTCTGCCCATCACTGAGGGTCCAGCGGGCTGAACCCATGTTCACTG 231
Db 37 Cys-----37
QY 232 CAGTCACCAACTCAGTCTCTGCTCTGACTATGACAGTAATCCACCAGCTCACTATG 291
Db 37 -----37
QY 292 GTGTGCAGTTACTGATGTGACCATGATGGGACTTTGAGATCGTCTGCGCGGGTACA 351
Db 38 -----ArgTrpLeuAlaIleCysArgLeuArgSerVal-----48
QY 352 ATGACCCCAACCTGGTTCTGAAGTATGACCGGCGCCAGAACGGCTGGTGAACATCGGG 411
Db 49 -----ProArgSerGly-----52
QY 412 TCGATGAGCGCAGCTCACTCTACTACGCGCTCGGGACCGGC-----AGGGGAACG 462
Db 53 -----ArgAlaSerArgSerThrTrpCysThrAlaProProValSerThrThr 70
QY 463 CCATTGGGTACAGCTCGACATCGACCGGAGCG-----GCCGGAGGAGA 510
Db 71 ArgLeuAlaThrValProProThrSerThrAlaCysAlaMetCysTrpAlaGlyCysArg 90
QY 511 TCTACTCTCTCAACACA-----ATAATGCTTCTCGGGGTGCCACGTACA 558
Db 91 ProAlaAlaSerGlyProGlyAlaCysCysSerSerProAlaProGlySerThrArgArg 110
QY 559 CCGACAAGTTGTTCAAGTTCC-----GCAATAACCGGTGGG 594
Db 111 ProMetAlaAlaGlySerThrArgSerArgArgPheProArgSerIleProAspGly 130
QY 595 AAGACATCTGAGCGATGAGTCACTGCGCCGCTGGTGTGGCCAGCCTCTTTGCCGAG 654
Db 131 SerCysSerThrProSerArgSerArgTrpThrAlaGlyTyr-----ArgLeuProAla 148
QY 655 GCTCTGTGCGCTGTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTATCATTTGCCAATT 714
Db 149 CysAlaTrpProAlaSerThrValAlaGly-----158
QY 715 ACGCCTACGGTANTGTGGGCCCTGATGCCCTCATTCGAATGACCCCTGAGGCCAGTGACC 774
Db 158 -----158
QY 775 TCTCCCGGGCAATCTCTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGGTCA---GCAAT 831
Db 159 -----GlySerGlyCysSer---163
```

```
QY 832 ATACAGGGCGGAGCGCTCAGCGTGGCGCCCATCTCAGCAGCAGTGCCTCGGATATCT 891
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 ThrArgCysValArgGlyThrGlyTrpSerAlaSerArgProCysMetProThrGlySer 183
QY 892 TCTGGCACAATGAGATGGCGCTAACTCTCTTTTCCACAAACGGGGCGATGGCACCTTTG 951
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 ThrProThrMetArgProGly-----CysTrpProSerCysCysAlaProMet 199
QY 952 TGGAGCTGCGGCCAGTGTGTGTGGAGCAGCCACCCACAGCATGGCGAGGTGTCGCC 1011
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 ProAlaAlaArg-----TrpArgThrAlaThrSerValSerThrThrArgPro 216
QY 1012 TGG-----CTGACTTCAACCGGTGATGGCAAGTGACATCGTCTATGGCAACT 1059
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
217 TrpArgCysThrArgTrpSerThrThrCysAlaSerAlaTrpAlaSer-----AlaAsn 234
QY 1060 GGA-----ATGCCCCCCACCGCTCTCTATCTGCAAATGAGCACCCCATG 1101
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
235 GlyProThrAsnIleArgTyrAlaAlaProAlaAlaSerValAlaAlaThrAlaAlaPro 254
QY 1102 GGAAGTCCGCT-----TCC 1116
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 GlyArgSerAlaGlyCysArgAlaIleProAlaThrAlaAlaArgAlaThrArgCysSer 274
QY 1117 GGCATATCGCCT-----CACCCAAGT 1137
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 GlyThrProAspGlyCysGlnProValValGlyAlaSerLeuAlaAlaTrpArgGlyAla 294
QY 1138 TCTCATGCT-----CTTCCC-----CTGTCGCA 1161
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 ValProCysGlyLeuLeuArgProGlyValArgProProGlnProPheGlyLeuArgHis 314
QY 1162 CGGTATACCCCGCACTTGTGACATGACAGGCTGGAGATCTTCTCAACACATG 1221
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 ArgLeuAspProPro-----ArgArgArgArgLeuProVal 326
QY 1222 CTTACCGAGCTCTCAGCAACCGCTCTCCGCGTCTATCCGTAGAGCAGCGAGACC 1281
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 ProArgAlaAla-----ProCysAlaAlaGly-ArgHisPro-----ValAlaAspG 342
QY 1282 CCTCATCGAGGAGCTCAATCCCGCGACGCTTGAGGCTGAGGCGCGGGCGACAGGG 1341
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
342 uProGlyArgAlaThrTyrArgAlaArgArgileArgAla-----AlaG 357
QY 1342 GTGTGTGACCCACTCGACGG-----AGACGGATGCTGGACCTCATCTTCT 1389
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 yLeuProAspAlaLeuCysArgGlyAsnAlaProAlaArgProAlaArgProGlnThrVa 377
QY 1390 CCTATGGAGAGTCCATGGCTCAGCGCTGCTCGCT-----CTTCCGG 1431
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
377 lAlaAlaArgLeuArgArgThrGluProGlyArgArgArgGlnArgGlyGlyTrpProG 397
QY 1432 GCAATCAGGG-----CTTCAACAACACTGGCTGCGAGTGGTCCACGACCC 1479
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
397 yValAlaAlaGlyArgGlyPheProAlaGlnArgArgile-----GlyValGlyLysPr 414
QY 1480 GGTGGGCGCTTGGCCAGGGAGC-----TAAGGTGCTGCTCTACACCAAGAGAGTG 1533
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
414 oProAlaGlyLeuAlaArgThrGlyLeuArgAlaSerArgArgLeuAlaProGlyGlyAr 434
QY 1534 GGGCCCACT-----GAGGATCATCGAGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
434 gGlyAlaProArgArgArgAsnHisArgThrProAlaArgLeuProGlySerAlaCysty 454
QY 1588 CGGTGGCACACTTGGCTGGGAGAGGATGAGCCAGCAGTGTGGA-----GGTACGT 1641
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
454 rArgGlyAla-----ArgSerGlyProThrGlyAlaVa 465
QY 1642 GGCAGATGCAAGATGGTGGCGGAGACGTGGC-----CAGCGGG 1683
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
465 lAlaAlaAlaProProAlaProProAspArgGlyPheProAlaCysLeuArgArgAlaAl 485
QY 1684 AGATGAACCTAGTGTGGAGATCCTTACCCCGGAGTGGAGACACACTTCAGGACCCAG 1743
```

```
Db 485 aAlaTyrLeuProThrAlaMetProGlyAlaProGlyGlnAlaProValAlaGlyArgTh 505
QY 1744 CCCACTGGA-----GTGTGCCCAAGGATTTCCCGCAGGAAATGCCATTGCA 1794
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 rAlaAlaGlyGlyThrArgProAlaProArgLeuLeuArgProGlyAlaLeu-----522
QY 1795 TGGACACCAATGAATGCATCCAGTTCCCATT-----CGTGTGCC 1833
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
523 -----LeuHisAspLeuProAlaPheHisArgGlyHisSerGlyAlaValPr 538
QY 1834 CTCGAGACAGCCCGTATGTCAACACTATGTGAAGTCTACAGGTCCGCGACCAAGA 1893
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
538 oAlaGlnArg-----GlnAspLeuSerPhePheGlnTyrGlyAlaArgAl 554
QY 1894 AGTCAGTCG-----GGCTACGAGCCCAACGAGATGGCACGCTCGTGGCTGTGGA 1950
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
554 aAlaGlnThrGlyGlyLeuArgValSerArgGlyLeuHisValArgArg-----ValAr 572
QY 1951 GGCCTGT 1957
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
572 gProCys 574

RESULT 13
US-09-252-991A-16655
; Sequence 16655, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16655
; LENGTH: 977
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16655

Alignment Scores:
Pred. No.: 8,31e-08 Length: 977
Score: 210.00 Matches: 226
Percent Similarity: 28.05% Conservative: 36
Best Local Similarity: 24.20% Mismatches: 253
Query Match: 5.08% Indels: 420
DB: 4 Gaps: 51

US-09-914-958B-35 (1-2177) x US-09-252-991A-16655 (1-977)
QY 1845 GCTTGTCTCGAGGGCACACGAATGGAGTGGATTCATTGTTCCATGCAATGCC 1786
Db 121 AlaArgLeuAlaGlyGlnArgGlyAlaThrAlaAlaAla-----133
QY 1785 CATTTCCTGTCTGGGAGAAATCCTTGGCCACACTCCAGTGGGGCTG-----GGTCTGAA 1732
Db 134 -----ProAlaArg-----ThrProGlyGlyLeuProArgGlyAlaPro 147
QY 1731 GTGTGTCCTCATCCCGGGGTAGAGGATCTCCAGGACTGAGTTTCATCTCCCGCTGGCCA 1672
Db 148 AlaArgProArgProGlyHisArg-----ValAlaValArgAlaGlyGlnArgSerArg 165
QY 1671 CGTTCGGCTCACCATTCTTGGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCT 1612
Db 166 ArgThrArgProPro-----ArgGlyProArgCysLeu-----Pro 177
QY 1611 TCCCCAGGC-----CAAAGTGTGCCCGGCTCCA 1582
```





```
Db 358 roSerCysThrAlaValAlaAla-TripAlaProAlaArgGlyGlyTyrAlaAlaAlaPro 377
QY 753 TTTCATCAGGGCATCAGGCCCCACATTACCGTAGGCGTAATTGGCAATGTAGATAGAGT 694
Db 378 ArgAsnSerCysSerAlaGly--ArgLeuArgArgArg-ValArgSerAlaArgAr 396
QY 693 AGGCTCCAGAGCCCTTCTCTCCACACAGGCCACAGAG-----656
Db 396 gAlaTrrArg-ProTrrProAlaGlyArgAlaThrProAlaSerAspAlaArgAsps 416
QY 655 -----CGTCCGGCAAGAGGCTGGCCACACACCGGCCACCGTTCACCTCAT 610
Db 416 exAlaSerArgGlyArgProAlaThrAlaAlaAspHisProAlaAlaTrrValSerAlaA 436
QY 609 CGCTCAGGATGCTTCCACCGGTTATTGC-----GGAACCTGAAACAACCTTGT 562
Db 436 laArgArgThrSerSerAlaProIle-AlaGlySerAlaProGlySerGlyThrAlaPro 455
QY 561 CGGTGT-----556
Db 456 ArgCysHisProValArgLysAspGlyAlaGlyProAlaSerThrGlyArgSerArgArg 475
QY 556 -----556
Db 476 ArgTrrAlaAlaAlaProAlaArgAlaArgThrGlyIleArgArgThrAlaAlaAla 495
QY 555 ---ACGTGGCCACCCCGGAGAGGCATTATTGTTGGGAAGTAGATCTCTCCCGGC 499
Db 496 AlaCysTyrProProProAlaPro-----GlyArgArgThrProAlaAla 510
QY 498 CPTCCCGCTCGATGTCCGAGGCTGTGACCCCATGGCTCCCTCCCGTCCCGCAGCG 439
Db 511 Arg--ArgArgSerAlaArgThrAlaProArg-----ProAlaIlePro-----524
QY 438 CGTAGTAGGTGAGTGCCTCATCGACCGCGATGTTCA-----400
Db 525 -----AlaHisArgProSerAlaSerAlaAlaThrGlyTrrProGlyPro 538
QY 399 -----CCAGCCGCTCTGGG-----CCGGTCATATCTTCAGAACCA 364
Db 539 AlaLeuAlaCysProAlaAlaAlaGlyArgProSerProAlaProAlaProAlaPro 558
QY 363 GGTGGGTCCATTGTACCCCG-----CCACGACGA-----TCTCAAAGTCCCATCAT 316
Db 559 Ala---ProHisAlaArgProAlaProAlaProArgArgProGlyArgSerProAlaThr 577
QY 315 GGTCCACATCAGTAATGTCACACCATAGTTGAGTGGTGGGATTACTGTGCATATCAG 256
Db 578 AlaProAla-----580
QY 255 GAGGCAGAACTGAGTTGGTACTGCAGTGAACATGGGTTACGCCCGCTGG---ACCCCT 199
Db 581 -----AlaPheAlaAlaProAlaGlyProThrPro 590
QY 198 CAGTGATGGGCAAGAAACACAGACGACGACGAGGAACGGTAAACATCTGGACATCCCGGGT 139
Db 591 -----AlaArgProAlaAlaAla-----ThrGly 599
QY 138 CAGCGCTCGAGCCCATCTCCCGCTCTCGGCCCGCCGCTAGGGCGGTGGGAAGCGGC 79
Db 600 ArgProSerAlaProProArgProAlaArgArgValArgTrrProTrrArgSer---618
QY 78 GTCGCTCGCGCTCTGCGCGCGCGCGCTGCTGCTGCTCCAGACCCCGGTCCCGGG 19
Db 619 AlaArgCys-----SerProArgArgProThrProArgSerAlaProGly 633
QY 18 CTGGCTCGAGCTCCCG 1
Db 634 AlaAlaThrGlyAlaPro 639
```

RESULT 15

US-09-252-991A-21920

; Sequence 21920, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21920
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-21920
```

## Alignment Scores:

Pred. No.:	2,43e-07	Length:	822
Score:	203.50	Matches:	177
Percent Similarity:	30.40%	Conservative:	34
Best Local Similarity:	25.50%	Mismatches:	204
Query Match:	4.92%	Indels:	279
DB:	4	Gaps:	40

US-09-914-958B-35 (1-2177) x US-09-252-991A-21920 (1-822)

```
QY 1719 CCGGGGGTAGAGGATCTCCAGCACTGAGTTCA---TCTCCCGCTGGCCACGTT-----1669
Db 27 ProGlyGlnArgArgArgProAlaAlaAlaGlnArgGlnArgGlnProArgArgGln 46
QY 1668 TCCGGCTCACCATCTTGC-----1651
Db 47 AlaGlyGlnProGlyAlaAlaGlyAspHisArgGlnProArgGlnGlnArgLeuArg 66
QY 1650 CATTTGGCCACCTCACTCCACACTGCTGGCTTCATCTTCCCGAGGCCAAAGTGTGCCA 1591
Db 67 HisProArgGlnProGlyArgHis-----GluProAlaProGlyArgArg-gPro 83
QY 1590 CCGGCTCCATCTCACACAGGTAGCTGAGCCCGCTCGATGATCTCAGGTGGGCCCCAC 1531
Db 84 AlaGlnProAlaGlyArgProAspLeuGlnProGlyArgArgProArgGlyAlaGlyArg 103
QY 1530 TCTTCTTGGTGTAGACACGACCTTAGCTCCCTGGCAAGGCCCAACCCGGTGGCTG 1471
Db 104 Gln-----ProGlyGlnProGlyHis-----111
QY 1470 GCACCACTCGCAGCCAGTTGTTTGAAGCCCTGATTGCCCGAAGACGACAGCGGCT 1411
Db 112 AlaPro-----GlyProGlyArgGlnProAlaAla 121
QY 1410 GAGCCATGGACTCTCCATGGGACAGATGAGTCCAGCATCCCTGCTC-----CGTCA 1357
Db 122 TyrArgTrrp-----ArgAlaGlyGlnProGlyArgProGlyGlnPro 136
QY 1356 AGTCGGTCACACACCCCTGTGCCCCCTCAGGCTCCAAAGCGCTCGCCGGATTCA 1297
Db 137 GlyArgGlnProArgProAlaGluArgGlnProArgGlnArgArgArg-----154
QY 1296 GCTCTCGATGAGGGGCTCTC-----CGTGTCTCTACGGATGACGCGGA 1252
Db 155 AlaGlnGlnArgGlnGlyLeuAlaAspAlaGlyHisArgAlaValArgGlnGlnArg---173
QY 1251 AGAGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGTGAAGAAGATCTCCAGCTCT 1192
Db 174 -----ArgArgHisProGlyAla 179
QY 1191 GGTCAATTCTCAAGTCGCGGTTGATGACCGTCCGACAGGGGAGGCATGGAACTTGG 1132
Db 180 ValAlaGlyAspSerArg-----186
```

```
QY 1131 GTGAGCGATGTCCTCCGGAAGCGGACCTTCCATGGGTGCTCATTTGCAGATAGAGCGGT 1072
Db      |||||
187 AlaArgArgAlaGlnProAlaGlyProSerLeuGlyAlaGly----- 200
QY 1071 GGGGGCCATTCCAGTTGCTAGCAGATGTCCTTTGCCATCATCGGTTGAAGTCAGCCA 1012
Db      |||||
201 -----ArgArgGlnProHisArgHisArgLeuArgGlnPro 213
QY 1011 GGGCGACACCTCGCCCATGCTGGTGGGGTGTCTCCACACAGCAGCTGCCCGCCAGCTCCA 952
Db      |||||
214 -----GlyTrpArgProLeuArgGlnArgProAlaGlnPro 225
QY 951 CAAGGTGCCATCCGCCGGTTGTGAAAGAAAGTTAGGCCCATCTCATTTGCGCAGA 892
Db      |||||
226 ArgArgProAlaPheProGlnProGlyArgGlyGlySerGlyArgGlnGlyArg 245
QY 891 AGATATCCGAGCCATGCTGCTGAGGATGGGGCCACCGCTGACGCTCGGCCCTGAT 832
Db      |||||
246 ArgAlaHisArg-----LeuGlnProGlyArgAlaGly----- 257
QY 831 ATTTGCTGACCCAGCTCAGCAGCCATCTCTGAGCGCCAGATGCCCGGAGAGGT 772
Db      |||||
258 -----GlnProLeuArgPro-----ValGlyLysArg 266
QY 771 CACTGGCCTCAGGTTCCATTTCAATGAGGCGCATCAGGCGCCACATTTACCGTAGCGTAA 712
Db      |||||
267 Lys-----ArgAlaAlaProAlaArgArgAspArg 277
QY 711 TGGCAATGTAGATAGTAGGTCCAGAGCCCTTTCTGTCCACAC----- 667
Db      |||||
278 GlnGlnArgArgGlnProAlaArgProArgProGlnArgGlnHisAlaAlaGlyArgTrp 297
QY 666 AGGCCA---CAGAGCGTCGGCAAGAGCGTGGCCACACCAGCGGCCAGCTGACCTCAT 610
Db      |||||
298 ArgProGluGlnArgLeuArgArgAlaGlyLysArgGlnProGlyProArg----- 314
QY 609 CGCTCAGGATGCTTCCCGACGGTTATTGCGGAACCTTGAAACACTTGCGGTGTACGTGG 550
Db      |||||
315 -----ProAlaThrGlyGlnProGly-----GlnArgArgAlaHis 327
QY 549 CCACCCCGGAGAGCATTTATGTTGAGGAAGTAGATCTCTCCCGGCGCTCCCGGT 490
Db      |||||
328 ProProHisArgGlnTrpHisLeuArgProGlyPheArgAla-----GlyAspProArg 345
QY 489 CGATCTCGCAGCTGTGACCCCAATGGGTTCCCTGCGCGTCCCGCAGCGGTAGTAGG 430
Db      |||||
346 Arg---ArgArgThrAspHisGlnTrp-----ProAlaGly---HisProArgGlnArg 361
QY 429 GTGAGCTGGCTCATCGACCGCGATGTTCAACGACGCGCTTCTGGGCGCGCTCATCTTCA 370
Db      |||||
362 MetAspGlnGlnArg-----AlaAlaSer-----ArgThrPro 373
QY 369 GAACACAGGTTGGGTCCATTGTATCCCGCCACACGATCTCAAAGTCCCGCAT----- 319
Db      |||||
374 GluProGlyHisArgHisLeuProProAspGlyArgGlyGlnAlaAlaGlyGlyAlaVal 393
QY 318 ---CATGTCCACATCAGTAAGTCCACACCATAGTTGAGCTGGG----- 277
Db      |||||
394 LeuHisTrpProArgArgLeuGluGlnArgArgProAlaGlyGlnArgArgGlnLeu 413
QY 276 -----TGGGAT----- 271
Db      |||||
414 AlaProArgProGluArgArgLeuProTrpGlnArgProArgHisGlnProArgArgLeu 433
QY 270 -----TAC 268
Db      |||||
434 ArgProGluArgArgGlnProArgProArgGlnArgGlnProArgArgArgGln 453
QY 267 TGTCTAGTCAGGAGCAGAACTGAGTTGGTGAAGTCAACATGGGTTCCAGCCCGCT 208
Db      |||||
454 CysHisAlaArgArg-----GlnProAla 462
QY 207 GGGACCCCTCAGTGTATGGGCGAGAAACAGAGCAGCAGCAGCAACCGGTAAACATCCTGGACA 148
```

```
Db      |||||
463 GlyGlnPro-----TrpAlaAspHisArgArgArgPro----- 474
QY 147 TGCCGGGGTCAGCGCTCGGAG-----CCATCTCCCGCTCTCGGCGCC 106
Db      |||||
475 ---ArgGlyGlnArgArgGluSerGluGlnLeuArgHisProGlyArgArgGlnPro 493
QY 105 -----CGCGCGCTAGGGCGGTGG----- 88
Db      |||||
494 AlaIleGluArgAlaArgProAlaGlnArgAlaArgValAlaValGlnTrpArgArgHis 513
QY 87 -----GAAGCG 82
Db      |||||
514 AspProAlaArgArgArgHisGlnProLeuArgGlyCysValGlnProArgArgHis 533
QY 81 GGCGCTCGCTCCCGCTCTCGCGCGCGCGCGCTGTGTCTCCAGCGCC----- 29
Db      |||||
534 GlyTyrArgProArgArgCysArgGln--Pro-----CysArgGlnProAlaGlnP 550
QY 28 -----CGGTCCCGGGCTGGCTCGAGCCT 5
Db      |||||
550 roPheArgGlyAspArgGluArgGlnGlyLeuGlnPro 562
```

Search completed: August 3, 2004, 19:50:34  
Job time : 67.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 3, 2004, 19:48:01 ; Search time 108 Seconds

(without alignments)  
12646.069 Million cell updates/sec

Title: US-09-914-958B-35

Perfect score: 4056

Sequence: 1 cggagagctcgagccagcc.....tctcttgaaaaa 2177

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 2582470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US09914958/runat\_03082004\_161654\_29816/app.query.fasta\_1.2375  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-TRANSL=0 -LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09914958 @cgn\_1\_1\_15 @runat\_03082004\_161654\_29816  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pcp.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pcp.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pcp.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pcp.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pcp.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
-----					

1	3292	81.2	618	12	US-10-332-929A-20
2	3190.5	78.7	653	12	US-10-332-929A-2
3	3190.5	78.7	653	15	US-10-295-027-183
4	3190.5	78.7	653	16	US-10-188-832-58
5	3181.5	78.4	637	12	US-10-332-929A-19
6	3179	78.4	612	12	US-10-332-929A-21
7	2337	57.6	1597	15	US-10-295-027-185
8	2337	57.6	1597	16	US-10-188-832-60
9	1967.5	48.5	418	15	US-10-295-027-181
10	1967.5	48.5	418	16	US-10-188-832-56
11	1492	36.8	306	10	US-09-892-877-247
12	1492	36.8	306	10	US-09-948-783-249
13	318	7.7	19662	15	US-10-084-846A-6
14	310	7.6	19723	15	US-10-084-846A-5
15	291	7.0	19723	15	US-10-084-846A-5
16	286.5	6.9	19725	15	US-10-084-846A-4
17	284	6.9	19608	15	US-10-084-846A-8
18	283	7.0	19608	15	US-10-084-846A-8
19	276.5	6.8	437	16	US-10-437-963-132079
20	273	6.6	19695	15	US-10-084-846A-3
21	272.5	6.7	19695	15	US-10-084-846A-3
22	267	6.6	50	10	US-09-948-783-125
23	267	6.6	51	10	US-09-892-877-124
24	259	6.3	19652	15	US-10-084-846A-7
25	257	6.3	1367	9	US-09-801-368-108
26	255	6.3	19725	15	US-10-084-846A-4
27	254	6.3	427	16	US-10-437-963-199493
28	253.5	6.2	5179	9	US-09-922-217-1068
29	253.5	6.2	5179	9	US-09-833-263-1068
30	253.5	6.2	5179	13	US-10-025-380-1068
31	250.5	6.2	19662	15	US-10-084-846A-6
32	249	6.1	528	10	US-09-840-746-20
33	244	6.0	19652	15	US-10-084-846A-7
34	243.5	6.0	497	16	US-10-437-963-125004
35	238.5	5.9	487	16	US-10-437-963-137896
36	237.5	5.9	377	16	US-10-437-963-133141
37	229.5	5.7	738	9	US-09-978-979-6
38	229.5	5.7	738	14	US-10-057-487-6
39	224	5.5	1259	14	US-10-260-715-8
40	223	5.5	417	16	US-10-437-963-143835
41	215.5	5.3	613	15	US-10-260-937-16
42	215.5	5.3	2080	15	US-10-353-690-36
43	213	5.3	465	16	US-10-437-963-174113
44	212	5.2	36	10	US-09-892-877-255
45	212	5.2	36	10	US-09-948-783-257

#### ALIGNMENTS

RESULT 1  
US-10-332-929A-20  
; Sequence 20, Application US/10332929A  
; Publication No. US20040072286A1  
; GENERAL INFORMATION:  
; APPLICANT: Akzo Nobel N.V.  
; TITLE OF INVENTION: Novel extracellular matrix protein  
; FILE REFERENCE: 2000.564  
; CURRENT APPLICATION NUMBER: US/10/332,929A  
; CURRENT FILING DATE: 2003-10-08  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 20  
; LENGTH: 618  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-332-929A-20  
Alignment Scores:  
Pred. No.: 8.57e-209 Length: 618  
Score: 3292.00 Matches: 618  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 81.16% Indels: 0



## Alignment Scores:

Pred. No.: 4,35e-202 Length: 653  
 Score: 3130.50 Matches: 603  
 Percent Similarity: 97.90% Conservative: 2  
 Best Local Similarity: 97.57% Mismatches: 2  
 Query Match: 78.66% Indels: 11  
 DB: 12 Gaps: 1

US-09-914-958B-35 (1-2177) x US-10-332-929A-2 (1-653)

QY 147 ATCTCAGGATGTTACCGTTCTGCTGCTCTGTTCTGCCATCACTAGGGGTCC 205  
 Db 1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20  
 QY 207 CACGGGCTGAACCCATGTTCTACTGCAGTCCACCAACTCAGTTCTGCTCTGCTGATGAC 266  
 Db 21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40  
 QY 267 AGTAATCCGACCCAGCTCAACTATGTTGGTGGAGTCTACTGATGGACCAATGATGGGAC 326  
 Db 41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60  
 QY 327 TTTGAGATCTCTGGGGGTACAATGGACCCCAACCTGTTCTGAAGTATGACCGGGCC 386  
 Db 61 PheGluLeuValAlaGlyTyrAsnGlyProAsnLeuValLeuIleTyrAspArgAla 80  
 QY 387 CAGAAGCGGTGGTGAACATCGCGGTGATGAGCGCAGCTCACCTACTACGCGTCCGG 446  
 Db 81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerProTyrTyrAlaLeuArg 100  
 QY 447 CACCGGAGGGGACGCATCTGGGTGACAGCTCGACATCGACGGGGAGCGCGGGAG 506  
 Db 101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120  
 QY 507 GAGATCTACTCTCAACACCAATAATGCTCTCGGGGTGGCCACGTACACCGACAAG 566  
 Db 121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140  
 QY 567 TTGTTCAAGTTCGGCAATAACCGGTGGGAGACATCTCGAGCGATGAGTCAACGTGGCC 626  
 Db 141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160  
 QY 627 CGTGGTGGCCAGCTCTTTGGCGGACGCTGTGTGGCCCTGTGTGGACAGAAAGGGCTCT 686  
 Db 161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180  
 QY 687 GGACGCTACTCTATCACTTGGCAATTACGCTACGGTAATGTGGGCCCTGATGCCCTC 746  
 Db 181 GlyArgTyrSerIleIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200  
 QY 747 ATTGAATGGACCTGAGCGCAGTGAACCTCTCCGGGGGCATTTCTGGCGCTCAGAGATGTG 806  
 Db 201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220  
 QY 807 GCTGCTGAGCTGGGTACACAAATATACAGGGGGCCAGCGCTACGCTGGGGCCCATC 866  
 Db 221 AlaAlaGluAlaGlyValSerLysTyrThrGlyArgGlyValSerValGlyProIle 240  
 QY 867 CTCAGCAGCAGTGCCTCGGATATCTCTCGACAAATGAGAAATGGCCCTAACTCTCTTTTC 926  
 Db 241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260  
 QY 927 CACAACCGGGCGATGGCACCTTTTGTGGACGCTCGCGCCAGTGTGTTGTGGACACCCC 986  
 Db 261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspPro 280  
 QY 987 CACAGCATGGCGAGTGTGCGCCCTGGCTGACTTCAACCGTGTATGCGCAAGTGGACATC 1046  
 Db 281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300  
 QY 1047 GTCTATGGCACTGGATGGCCCCCAGCGCTCTATCTGCAATGACACCCCATGGGAAG 1106  
 Db 301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

QY 1107 GTCCGCTTCGGGACATCGCCTCACCAAGTTCTCCATGCCCTCCCTGTCGCGACCGCTC 1166  
 Db 321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340  
 QY 1167 ATCACCGCGACATTTGACATGACCGAGCTCGAGATCTTCTTCAACAACATTTGCCATC 1226  
 Db 341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePheAsnAsnIleAlaTyr 360  
 QY 1227 CGCAGCTCTCTCACCAACCGCCTCTTCGCGCTCATCGTAGAGACGACGAGACCCCTC 1286  
 Db 361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380  
 QY 1287 ATCAGAGAGCTCAATCCCGCGACGCTTGGAGCCCTGAGGCGCGGGSCACAGGGGGTGT 1346  
 Db 381 IleGluLeuLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyVal 400  
 QY 1347 GTACCGACTTCACCGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGTCCATG 1406  
 Db 401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420  
 QY 1407 GCTCAGCGCTGTCCGCTCTTCGCGGCAATCAGGGCTTCAACAACACTCGCTCGCAGTG 1466  
 Db 421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440  
 QY 1467 GTCCACCGCACCCGTTTGGGCTTTGCGAGGGAGCTTAAGTCTGTCTCACCAAG 1526  
 Db 441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460  
 QY 1527 AAGAGTGGGGCCACCTGAGGATCATCGAGGGGGCTCAGGCTACCTGTCTGAGATCGAG 1586  
 Db 461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480  
 QY 1587 CCGCTGGCACACTTGGCTTGGGAAAGATGAAAGCCAGCAGTGTGGAGGTGACCTGGCCCA 1646  
 Db 481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500  
 QY 1647 GATGGCAAGATGTGACCGGAACTGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATC 1706  
 Db 501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520  
 QY 1707 CTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCTGAGTGTGGCCAGGA 1766  
 Db 521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540  
 QY 1767 TTCTCCAGCAGCAAAATGGCCATTGTCATGGACACCAATGAATGCATCCAGTTCCTCATC 1826  
 Db 541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560  
 QY 1827 GTGTGCCCTCGACAAAGCCCGTATGTGTCAACACCTATGGAAAGCTACAGTGCCTGGG 1886  
 Db 561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580  
 QY 1887 AACAGAAAGTCACTCGGGGTACGAGCCCAACGAGGATGGCAGCCTCGCTGGGCTGG 1946  
 Db 581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599  
 QY 1947 TGGAGCCCTGTGTGAAGATAGTAGACACCAACAAAGTTGGGAAGAGCTTGTGCC 2000  
 Db 600 -----ThrLeuGlyGlnSerProGlyPro 607

## RESULT 3

US-10-295-027-183  
 ; Sequence 183, Application US/10295027  
 ; Publication No. US20030232350A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Afar, Daniel  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Ginsberg, Wendy M.  
 ; APPLICANT: Gish, Kurt C.  
 ; APPLICANT: Glynn, Richard  
 ; APPLICANT: Hevezi, Peter A.  
 ; APPLICANT: Mack, David H.





```
QY 1347 GTGACCGACTTCGACGGACGGGATGCTGGACCTCATCTTGCCTCCATGAGAGTCCATG 1406
Db ValThrAspPheAspGlyMetLeuAspLeuLeuLeuSerHisGlyGluSerMet 420
QY 1407 GCTACGCGCTCTCGCTTCTCCGGGGCAATCAGGGCTTCAACAACAACCTGGCTGGAGTG 1466
Db AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440
QY 1467 GTGCCACGACCGGTTTGGGCGCTTGGCAGGGAGCTAAGTCTGCTCTACACCAAG 1526
Db ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460
QY 1527 AAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGCTGAGATGGAG 1586
Db LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480
QY 1587 CCCGTGGCACACTTTGGCTGGGGAGGATGAAGCCAGCAGTGTGAGGTGACGTGGCCA 1646
Db ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500
QY 1647 GATGGCAAGATGGTGGCGGAACGTTGGCCAGCGGGAGATGAACCTAGTCTGGAGATC 1706
Db AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520
QY 1707 CTCTACCCCGGATGAGACACACTTCAGGACCCAGCCCACTGGAGTGTGSCCAAGGA 1766
Db LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540
QY 1767 TTCTCCACGAGAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560
QY 1827 GTGTGCCCTCGACAAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGTGTCCGGACC 1886
Db ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580
QY 1887 AACAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCTCGTGGGCTGG 1946
Db AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599
QY 1947 TGGAGCCCTGTGTGAAGATAGTACACCAAGTTGGGAAGACCTTGGTCCC 2000
Db :::::-----ThrLeuGlyGlnSerProGlyPro 607
```

## RESULT 5

```
US-10-332-929A-19
; Sequence 19, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: patentin version 3.2
; SEQ ID NO 19
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-332-929A-19
```

```
Alignment Scores:
Pred. No.: 1.7e-201 Length: 637
Score: 3181.50 Matches: 602
Percent Similarity: 97.89% Conservative: 2
Best Local Similarity: 97.57% Mismatches: 6
Query Match: 78.44% Indels: 7
DB: 12 Gaps: 1
```

US-09-914-958B-35 (1-2177) x US-10-332-929A-19 (1-637)

```
QY 147 ATGTCCAGGATGTACCGTTCTCTGCTGCTCTGTTCTTCTGCCATCACTAGGGGTCC 206
Db MetSerArgMetLeuProPheLeuLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20
QY 207 CAGCGGGGTGAACCCATGTTCTAGTCACTGAGTCAACCACTCAAGTCTGCTCTGACTATGAC 266
Db GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProAspTyrAsp 40
QY 267 AGTAATCCACCCAGCTCAACTATGTTGGCGAGTTACTGATGCTGACCATGATGGGAC 326
Db SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60
QY 327 TTTTGAATCGTCTGGCGGGGTACATGACACCACTCGTTCTTCTGAAGTATGACCGGGCC 386
Db PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80
QY 387 CAGAGCGGCTGGTGAAATCGCGGTGATGAGCGGAGCTCACCCTACTACGGCGTGGCG 446
Db GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100
QY 447 GACCGCGAGGGNAACGCCATTGGGTCACAGCTCGACATCGACGGGACGGCCCGGAG 506
Db AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120
QY 507 GAGATCTACTTCTCAACACCAATATGCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140
QY 567 TTGTTCAAGTTCGCGAATAACCGGTGGGAAGACATCTCTGAGCGATGAGTCAACGTGGCC 626
Db LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160
QY 627 CTGTGGTGGCCAGCGCTCTTTCGGGACGCTCTGTGGCGCTGTGTGACAGAAAGGCTCT 686
Db ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180
QY 687 GGACGCTACTCTATCTACATTGCCAATTACGCTACGGTAAATGTGGGCGCTGATGCCCTC 746
Db GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200
QY 747 ATTGAATGACCTGAGCGCCAGTGACCTCTCCCGGGCATTTCTGGGCTCAGAGATGTC 806
Db IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220
QY 807 GCTGCTGAGCTGGGCTCAGCAAAATACAGGGGGCGGAGGCGTCAAGCTGGGCGCCCATC 866
Db AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240
QY 867 CTCAGCAGCAGTGCCTCGGATATCTTCTCGCAATGAGAATGGGCGCTAACCTTCTTTTC 926
Db LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260
QY 927 CACAACGGGGGATGGCACCTTTTGGACGCTGGCGGCGAGTCTGCTGGTGGACGACCCC 986
Db HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280
QY 987 CACGACATGGGCGAGTGTGCGCCCTGGCTTCAACCGTGTATCGCAAGTGGACATC 1046
Db HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300
QY 1047 GTCTATGGCACTGGGAATGGCCCCCAGCGCTCTATCTGCAATGAGACCCATGGGAAG 1106
Db ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320
QY 1107 GTCCGCTTCGGGAGATCGCTCACCAAGTTCTCCATGCCCTCCCTCCCTCCGACGGTC 1166
Db ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340
QY 1167 ATCACCGCGCACTTTCACAATGACACGAGCTGGAGATCTTCTTCAACAACTTGGCTAC 1226
Db IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360
QY 1227 CGCAGCTCTCAGCCCAACCGCGCTCTTTCGCGCTCATCTCGTAGAGACGACGAGACCCCCTC 1286
```





1167	ATCACCCCGCAGTCTTGCATATGACACAGAGCTGGAGATCTTCTTCAACAACTGCTAC	1226	;	PRIOR APPLICATION NUMBER: US 09/663,733
341	1leThrAlaAspPheAsnAspGlnGluLeuGluPhePheAsnAsn1leAlaTyr	360	;	PRIOR FILING DATE: 2000-09-15
1227	CGACGCTCTCAGCAACCGCTCTTCGCGTCATCCGTCAGAGAGCAGCGAGACCCCTC	1286	;	PRIOR APPLICATION NUMBER: US 60/350,666
361	ArgSerSerAlaAsnArgLeuPheArgVal1leArgArgGluHisGlyAspProLeu	380	;	PRIOR FILING DATE: 2001-11-13
1287	ATCGAGGAGCTCAATCCCGGCACCGCTTGGAGCCTGAGGGCCGGGCGACAGGGGTGG	1346	;	PRIOR APPLICATION NUMBER: US 60/335,394
381	1leGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal	400	;	PRIOR FILING DATE: 2001-11-15
1347	GTGACCGACTTCAGCGAGACGGATGCTGGACCTCATCTTCCATCGGAGAGCCATG	1406	;	PRIOR APPLICATION NUMBER: US 60/332,464
401	ValThrAspPheAspGlyAspGlyMetLeuAspLeu1leLeuSerHisGlyGluSerMet	420	;	PRIOR FILING DATE: 2001-11-21
1407	GCTCAGCGCTGTCCGCTCTTCGCGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTG	1466	;	PRIOR APPLICATION NUMBER: US 60/334,393
421	AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnTrpLeuArgVal	440	;	PRIOR FILING DATE: 2001-11-29
1467	GTGCCACGACCCGCTTGGGCGCTTTCGAGGGAGCTAAGTCTGCTCTACACCAAG	1526	;	PRIOR APPLICATION NUMBER: US 60/340,376
441	ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys	460	;	PRIOR FILING DATE: 2001-12-14
1527	AAAGTGGGGCCCATCTGAGGATCATCGACGGGCTCAGGCTACCTGTGCTGAGATGG	1586	;	PRIOR APPLICATION NUMBER: US 60/347,211
461	LysSerGlyAlaHisLeuArg1le1leAspGlyGlySerGlyTyrLeuCysGluMetGlu	480	;	PRIOR FILING DATE: 2002-01-08
1587	CCCGTGGCACACTTTGGCTGGGAGAGATGAAGCCAGCAGTGTGGAGGTGCGGCCA	1646	;	PRIOR FILING DATE: 2002-01-10
481	ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro	500	;	PRIOR APPLICATION NUMBER: US 60/355,250
1647	GATGCAAGATGTGAGCGGAACGTGGCCAGCGGGAGATGAACCTCAGTCTGGAGATC	1706	;	PRIOR FILING DATE: 2002-02-08
501	AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGlu1le	520	;	PRIOR APPLICATION NUMBER: US 60/356,714
1707	CTCTACCCCGGATGAGACACACTTCAGGACCCAGCCCTCAGGCTGAGTGTGGCAAGGA	1766	;	PRIOR FILING DATE: 2002-02-13
521	LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly	540	;	Remaining Prior Application data removed - See File Wrapper or PALM.
1767	TTCTCCACGAGAAATGGCCATTCATGGACACCAATGAATGCATCCAGTTCCCATTC	1826	;	NUMBER OF SEQ ID NOS: 1386
541	PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCys1leGlnPheProPhe	560	;	SEQ ID NO 185
1827	GTGTGCTTCGACAGACCGCGTATGTGTCAACACCTATGGAAGCTACAGTGTCCCGGACC	1886	;	SOFTWARE: Patentin Ver. 2.1
561	ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr	580	;	LENGTH: 1597
1887	ACAAGAAGTGCAGTCGGGCTACGAGCCCAACGAGGATGGCACAGCTGCGTG	1940	;	TYPE: PRT
581	AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysVal	598	;	ORGANISM: Homo sapiens
RESULT 7				US-10-295-027-185
Sequence 185, Application US/10295027				US-09-914-958B-35 (1-2177) x US-10-295-027-185 (1-1597)
Publication No. US20030232350A1				QY 145 GCATCTCCAGGATGTACCGTTCCTGCTGCTGCTGGTTTCTGCCATCCTACGAGGGT 204
GENERAL INFORMATION:				Db 2 AlaCysProGlyGlyLeu---ProAlaArgCysSerGlyTrpMet---GlyLeuGlyGly 19
APPLICANT: Afar, Daniel				QY 205 CCCAGCGGCTGAACCCATGTTCACTGCAGTCAACCACTCAGTTCTGCCTCCTGACTATG 264
APPLICANT: Aziz, Natasha				Db 20 ProSerGlySerSerProAlaSer----- 27
APPLICANT: Ginsberg, Wendy M.				QY 265 ACAGTAATCCACCCAGCTCAACTATGTGTGGCAGTTACTGATGTGGACCATGATGGG 324
APPLICANT: Gish, Kurt C.				Db 28 -----ProPro----- 29
APPLICANT: Glynn, Richard				QY 325 ACTTTGAGATCGTCGTGGCGGGTACAATGGACCCCAACCTGTTCTGAAGTATGACCGGG 384
APPLICANT: Hevezi, Peter A.				Db 30 -----HisSerSerSerArg--TyrAsnGlyProAsnLeuValLeuLysTyrAspArg 47
APPLICANT: Mack, David H.				QY 385 CCCAGAAGCGGTGTGTGAACATCGCGTGCATGAGCGCAGCTCACCTACTACGCGTGC 444
APPLICANT: Murray, Richard				Db 47 laglnlysaargLeuValAsn1leAlaValAspGluArgSerSerProTyrTyrAlaLeuA 67
APPLICANT: Watson, Susan R.				QY 445 GGGACCGGAGGGAACGCCATTGGGGTACAGCCTCGACATCGAGCGGACCGCGGG 504
APPLICANT: Bos Biotechnology, Inc.				Db 67 rgAspArgGlnGlyAsnAla1leGlyValThrAlaCysAsp1leAspGlyAspGlyArgg 87
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and				QY 505 AGGAGATCTACTTCTCTCAACCAATAATGCCCTTC----- 539
METHODS OF SCREENING FOR MODULATORS OF CANCER				Db 87 luGlu1leTyrPheLeuAsnThrAsnAsnAlaPheSerGlyHisSerSerSerAlaGlnV 107
FILE REFERENCE: 018501-0125000S				QY 539 ----- 539
CURRENT APPLICATION NUMBER: US/10/295,027				Db 107 alProSerGlyLeuHisArgAsnArgProValLeuLysProProThrThrProAlaG 127
CURRENT FILING DATE: 2002-11-13				QY 539 ----- 539

Db 127 lyLeuLeuGlyLeuProLeuSerGlyArgAspPheSerSerSerLeuGlyGlnAla 147  
 QY 539 ----- 539  
 Db 147 erProAspSerArgGlnGlyGluArgValProValProCysCysArgGlyGlyLeuArgp 167  
 QY 540 -----TCGGGGTGGCCACGTACACCG 561  
 Db 167 roThrHisGluProGluProPheLeuLeuArgProLysSerGlyValAlaThrTyrThr 187  
 QY 562 ACAAGTTGTCAGTTCGCAATAACCGGTGGAGACATCTGAGCGATGAGGTCAACG 621  
 Db 187 spLysLeuPheLysPheArgAsnAsnArgTTrpGluAspIleLeuSerAspGluValAsnV 207  
 QY 622 TGGCCCGTGGTGGCCAGCCTCTTTCGGGACGCTCTGTGGCTGTGTGGACAAAGG 681  
 Db 207 alAlaArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysG 227  
 QY 682 GCTCTGGACGCTACTCTATCTACATTGCCAATTACGCTACGGTAATGTGGCCCTGATG 741  
 Db 227 lySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspA 247  
 QY 742 CCTCATGTAATGACCTGAGCCAGTCAGTACCTCTCCGGGGCATTTCTGGCGCTCAGAG 801  
 Db 247 laLeuIleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgA 267  
 QY 802 ATGTGGCTGCTGAGCTGGGTCTAGCAATATACA----- 836  
 Db 267 spValAlaAlaGluAlaGlyValSerLysTyrThrGluGlyPheSerHisThrAlaSerP 287  
 QY 836 ----- 836  
 Db 287 roSerIleGlyGluIleSerGlyArgThrGluGluArgGluGlyGlyAspProGluGluA 307  
 QY 836 ----- 836  
 Db 307 laAspGluGluHisSerGlyAspGlySerThrSerGlnLeuCysArgLeuGlyTrpLysA 327  
 QY 836 ----- 836  
 Db 327 spGlyGlnPheLysGluGluAlaAlaAlaLeuValGluGluGlnArgGluAlaGlyAlaA 347  
 QY 836 ----- 836  
 Db 347 laGlyValProArgGlyArgValArgThrAlaLeuGlnThrSerLysSerHisLeuAlaA 367  
 QY 836 ----- 836  
 Db 367 spLysAsnLeuPheGlyProProCysTyrTyrSerValCysAlaProSerProAlaHisP 387  
 QY 836 ----- 836  
 Db 387 roPheProAlaArgGlnAlaProGlnHisTyrProValAlaProLeuValThrGlnLeuM 407  
 QY 836 ----- 836  
 Db 407 etThrHisGlyArgLeuAlaGlyLysLeuAlaArgSerValProHisProArgAlaProG 427  
 QY 836 ----- 836  
 Db 427 lyMetAspProLysCysLysGlyArgHisAlaGluProGlyLeuMetAlaGluAlaLeuG 447  
 QY 836 ----- 836  
 Db 447 lyAlaTrpProAlaLeuSerThrThrValValProGlyGlyLeuArgSerTrpGluGluS 467  
 QY 836 ----- 836  
 Db 467 erArgGlnLysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTrpS 487  
 QY 836 ----- 836  
 Db 487 erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProIleL 507

QY 836 ----- 836  
 Db 507 euGlnArgThrAspGlyAspProGlyArgArgAspSerProLysValThrGlnGluC 527  
 QY 836 ----- 836  
 Db 527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547  
 QY 836 ----- 836  
 Db 547 ysArgGluIleGlyArgGluThrGlyAlaValAlaGlyArgProLeuSerHisProLeuValP 567  
 QY 836 ----- 836  
 Db 567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587  
 QY 836 ----- 836  
 Db 587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607  
 QY 836 ----- 836  
 Db 607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA 627  
 QY 836 ----- 836  
 Db 627 rgGluAlaGluPheProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647  
 QY 836 ----- 836  
 Db 647 euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlys 667  
 QY 837 -----GGGGCGCGAGCGCTCAGCGTGGCGCCCATCTCTCAGCAGCGTGCCT 882  
 Db 667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerSerAlaS 687  
 QY 883 CGGATATCTTCTCGACAATGAGATGGGCTTAACCTTCCTTTTCCAAACCGGGCGCATG 942  
 Db 687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707  
 QY 943 GCACCTTTGGAGCGCTGGCGCCAGTCT----- 971  
 Db 707 lyThrPheValAspAlaAlaAlaSerAlaGluArgLeuAlaPheIleValHisLeuL 727  
 QY 971 ----- 971  
 Db 727 ysTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS 747  
 QY 971 ----- 971  
 Db 747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisGlyL 767  
 QY 971 ----- 971  
 Db 767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787  
 QY 971 ----- 971  
 Db 787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807  
 QY 971 ----- 971  
 Db 807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL 827  
 QY 971 ----- 971  
 Db 827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847  
 QY 972 -----GGTGTGGACGACCCCGACATGGGCGAGGTGTTCGCCCTGGCTGACTTCAACC 1026  
 Db 847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867

QY 1027 GTGATGGCAAGTGGACATCGTCTATGCACTGAATGGCCCGCCACCGCCTCTATCGC 1086  
DB 867 rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887  
QY 1087 AAATGAGCACCCATGGGAAGTCCCGTCCGGGACATCGCCTCACCCAAAGTTCTCCATGC 1146  
DB 887 InMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSerMetP 907  
QY 1147 CTTCCCTGTCCGCGACGGTCAATCCCGGACATCGCCTCACCCAAAGTTCTCCATGC 1206  
DB 907 roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP 927  
QY 1207 TCTTCAACAACTTGCCTACCGCAGCTCTCCAGCAACCGCCTCTCCGC----- 1256  
DB 927 hePheAsnAsnIleAlaTyrArgSerSerAlaAsnArgLeuPheArgCysSerIleL 947  
QY 1256 ----- 1256  
DB 947 euAlaArgGlySerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA 967  
QY 1256 ----- 1256  
DB 967 rgIleArgArgGlyGlyPheProGlyProGlyGlnAlaLysValAsnThrGlyProL 987  
QY 1256 ----- 1256  
DB 987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007  
QY 1256 ----- 1256  
DB 1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysAsnValA 1027  
QY 1256 ----- 1256  
DB 1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047  
QY 1256 ----- 1256  
DB 1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067  
QY 1256 ----- 1256  
DB 1067 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1087  
QY 1256 ----- 1256  
DB 1087 rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107  
QY 1256 ----- 1256  
DB 1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS 1127  
QY 1256 ----- 1256  
DB 1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147  
QY 1256 ----- 1256  
DB 1147 roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167  
QY 1256 ----- 1256  
DB 1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187  
QY 1256 ----- 1256  
DB 1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207  
QY 1256 ----- 1256  
DB 1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227  
QY 1256 ----- 1256

DB 1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247  
QY 1256 ----- 1256  
DB 1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267  
QY 1256 ----- 1256  
DB 1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlnS 1287  
QY 1257 -----GTATCCCTAGAGACGACGAGACCCCC 1284  
DB 1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307  
QY 1285 TCATCGAGGAGCTCAATCCCGCGAGCCCTCGAGCCTCAGCGCCGGGCGACACAGGGGGTG 1344  
DB 1307 euIleGluLeuLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327  
QY 1345 TGGTGACCCGACTTCGACGAGAGCGGATGCTGACCTCATCTTGTCCCATGGAGAGTCCA 1404  
DB 1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347  
QY 1405 TGGCTCAGCCGCTGCTCCGCTTCCGGGCAATCAGGGCTTACACAACTGCTCGCGAG 1464  
DB 1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnTrpLeuArgV 1367  
QY 1465 TGGTGCCACGACCCCGTGGGCTTGGGCTTCCAGGGGAGCTAAGGTCGTCTCTACACCA 1524  
DB 1367 alValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrL 1387  
QY 1525 AGAAGAGTGGGGCCACCTGAGGATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGG 1584  
DB 1387 yslsSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetG 1407  
QY 1585 AGCCCTGGGCACACTTGGCTGGGGAAGATGAAGCCAGCAGTGTGGAGTGCACGTGC 1644  
DB 1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpP 1427  
QY 1645 CAGATGCGAAGTGTGTAGCGGAGCGGAACTGGCCAGCGGGGAGATGAACACTCAGTGTGGAGA 1704  
DB 1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447  
QY 1705 TCCTCTACCCCGCGGATGAGCACACACTTCAGACCCCGCCACTGGAGTGTGGCCAAAG 1764  
DB 1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467  
QY 1765 GATCTCCCGACGAGAAATGGCCATTGCATGACACCAATGAATGCATCCAGTTCCTCCAT 1824  
DB 1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProp 1487  
QY 1825 TCGTGTGCCCTCGAGACAAGCCGTATGTGTCAACACTATGGAAGCTACAGGTGCCGGA 1884  
DB 1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507  
QY 1885 CCAACAAAGAGTGCAGTCCGGGCTACGAGCCCAACAGGAGTGGACACAGCTCGGTGGGC- 1943  
DB 1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyT 1527  
QY 1944 -----TGGTGGAGCCCTGTGTTCAAG----- 1964  
DB 1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547  
QY 1965 -----ATAGTGACACCAACCAAG----- 1980  
DB 1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566  
QY 1981 TTGGGAGAGCCCTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGATA 2040  
DB 1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrThrLeuLeuGlnLysAlaProGlyIle 1586  
QY 2041 CCTGTTGATCAG 2052  
|||  
|||



Db	467	erArgGlnIysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTps	487
QY	836	-----	836
Db	487	erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProIleL	507
QY	836	-----	836
Db	507	euGlnArgThrAspGlyAspProGlyArgArgAspSerProIysValThrGlnGluC	527
QY	836	-----	836
Db	527	ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL	547
QY	836	-----	836
Db	547	ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP	567
QY	836	-----	836
Db	567	roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP	587
QY	836	-----	836
Db	587	roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG	607
QY	836	-----	836
Db	607	lulysGluGluGlyIysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA	627
QY	836	-----	836
Db	627	rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL	647
QY	836	-----	836
Db	647	euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlys	667
QY	837	-----GGGGCGGAGGCGTCAGCGTGGGCGCCCATCCCTCAGCAGCAGTGCCT	882
Db	667	erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerSerAlaS	687
QY	883	CGGATATCTTCGCAATGAGATGGGCTAACTTCCTTTTCCACACCGGGCGCATG	942
Db	687	erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG	707
QY	943	GCACCTTGTGGACGTCGGCCAGTCCT	971
Db	707	lyThrPheValAspAlaAlaAlaSerAlaGluArgLeuAlaPheIleValHisLeuL	727
QY	971	-----	971
Db	727	yeTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS	747
QY	971	-----	971
Db	747	erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisGlyL	767
QY	971	-----	971
Db	767	euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA	787
QY	971	-----	971
Db	787	laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL	807
QY	971	-----	971
Db	807	euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL	827
QY	971	-----	971
Db	827	eutrpSerAlaIleProGluSerLeuMetThrHisSertyrLeuSerSerGluArgValA	847
QY	972	-----GGTGTGACGACCCCGCCAGCATGGCGAGGTGTCCCTGGCTGACTTCAACC	1026
Db	847	snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA	867
QY	1027	GTGATGGCAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCCTCTATCTGC	1086
Db	867	rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG	887
QY	1087	AAATGAGCACCCTGGGAAGTCCGCTTCGGGACATCCGCTCACCACCAAGTTCTCCATGC	1146
Db	887	lnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSerMetP	907
QY	1147	CCTCCCTCTCCGCGGTCATCACCGCGACTTTGACAATGACACAGGAGCTGGAGATCT	1206
Db	907	roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP	927
QY	1207	TCTTCAACAACTGCTACCGCAGCTCTCAGCCCAACCGCTCTTCGCGC	1256
Db	927	hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL	947
QY	1256	-----	1256
Db	947	euAlaArgGlySerSerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA	967
QY	1256	-----	1256
Db	967	rgIleArgArgGlyGlyPheProGlyProGlyGlyGlnAlaLysValAsnThrGlyProL	987
QY	1256	-----	1256
Db	987	euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG	1007
QY	1256	-----	1256
Db	1007	lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA	1027
QY	1256	-----	1256
Db	1027	laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG	1047
QY	1256	-----	1256
Db	1047	lyLeuGlnGlyPheThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL	1067
QY	1256	-----	1256
Db	1067	ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA	1087
QY	1256	-----	1256
Db	1087	rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT	1107
QY	1256	-----	1256
Db	1107	yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS	1127
QY	1256	-----	1256
Db	1127	erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP	1147
QY	1256	-----	1256
Db	1147	roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG	1167
QY	1256	-----	1256
Db	1167	lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT	1187
QY	1256	-----	1256
Db	1187	yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG	1207

```
QY 1256 ----- 1256
Db 1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227
QY 1256 ----- 1256
Db 1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247
QY 1256 ----- 1256
Db 1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267
QY 1256 ----- 1256
Db 1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlnS 1287
QY 1257 -----GTCATCCGTAGAGACGACGAGACGAGACCCCC 1284
Db 1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307
QY 1285 TCATCGAGGAGCTCAATCCCGGACGCGCTTGAGAGCTGAGGCGCGGGCACAGGGGTG 1344
Db 1307 euIleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327
QY 1345 TGTGTACCGACTTCGACGAGAGCGGATGCTGGACCTCATCTGTCCTCCATGGAGATCCA 1404
Db 1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347
QY 1405 TGGCTCAGCGCGCTGCTCCGTCTCCGGGGCAATCAGGGCTTCAACAACTGGCTGGCGAG 1464
Db 1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgV 1367
QY 1465 TGTGTGCACACACCGGTTTGGGGCTTTGCCAGGGAGCTAAGTTCGTCTTACACCA 1524
Db 1367 alValProArgThrArgPheGlyAlaPheAlaargGlyAlaLysValLeuTyrThrL 1387
QY 1525 AGAAGAGTGGGGCCCACTCAGGATCATTCAGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
Db 1387 yLysSerGlyAlaHisLeuArgIleLeuAspGlyGlySerGlyTyrLeuCysGluMetG 1407
QY 1585 AGCCCTGGGCACACTTTGGCGTGGGGAAGGATGAAGCCAGGCTGGAGGTGACGTGGC 1644
Db 1407 luProValAlaHisPheGlyLeuGlyLysaspGluAlaSerSerValGluValThrTrpP 1427
QY 1645 CAGATGCAAGATGGTGAGCGGAACTGTCGCCAGCGGGAGATGAATCAGTCTCTGAGA 1704
Db 1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447
QY 1705 TCCTCTACCCCGGGATGAGGACACACTTCAGACCCGACCCCACTGGAGTGGGCAAG 1764
Db 1447 leuTyrProArgaspGluaspIhrLeuGlnaspProAlaProLeuGluCysGlyGlnG 1467
QY 1765 GATTCTCCAGCAGGAAATGGCCATTGTCATGCACACCAATGAATGATCCAGTCCCAT 1824
Db 1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProp 1487
QY 1825 TGTGTGCCCTCAGACACAGCCCGTATGTCTAACACCTATGGAAGCTACAGTGGCGGA 1884
Db 1487 heValCysProArgaspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507
QY 1885 CCAACAAGAGTCACTCGGGGTACGAGCCCAACGAGGATGACAGCCTGCTGGTGGC- 1943
Db 1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluaspGlyThrAlaCysValGlyT 1527
QY 1944 -----TGGTGGAGCCCTGTTTGAAG----- 1964
Db 1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547
QY 1965 -----ATATGACACACCAAG----- 1980
Db 1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566
QY 1981 TTGGGAAGACCTTGGTCCCTGATCACTCAATCACTGCTGCAATCAACCGCCTGGAATA 2040
```

```
Db 1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586
QY 2041 CCTGTTGATCAG 2052
Db 1587 ProGluAlaGln 1590

RESULT 9
US-10-295-027-181
; Sequence 181, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Heveri, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 181
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-181

Alignment Scores:
Pred. No.: 2,1e-121 Length: 418
Score: 1967.50 Matches: 385
Percent Similarity: 89.81% Conservative: 3
Best Local Similarity: 89.12% Mismatches: 10
Query Match: 48.51% Indels: 34
DB: 15 Gaps: 2

US-09-914-958B-35 (1-2177) x US-10-295-027-181 (1-418)
QY 753 ATGGACCTGAGCCAGTGAACCTCTCCGGGGCAATCTGGCGCTCAGATGGCTGCT 812
Db 1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20
QY 813 GAGGCTGGGTGAGCAATATACAGGGGCGGAGGCTCAGGCTGGGCCCATCTCAGC 872
Db 21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40
```

```
QY 873 AGCAGTGGCTGGATATCTTTCGACAAATGAGATGGCCCTAACTTCTCTTTTCCACAC 932
Db |||||
QY 41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60
Db |||||
QY 933 CGGGCGATGGACCTTTGTGACGCTGGCGCAGTGTGTGGTGGACACCCACCCACAG 992
Db |||||
QY 61 ArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspProHisGln 80
Db |||||
QY 993 CATGGCGAGGTGTGCGCTGCTGATTCACCGGTGATGGCAAAAGTGGACATCTCTAT 1052
Db |||||
QY 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100
Db |||||
QY 1053 GGCACATGGAGTGGCCCGCCACCGCTCTATCTGCAATGAGCCATGGCAAGTGGACATCTCTAT 1052
Db |||||
QY 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120
Db |||||
QY 1113 TTCCGGGACATCGCTCACCCCAAGTTCTCATGCGCTCCCTCGCACGGTCTCATCACC 1172
Db |||||
QY 121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140
Db |||||
QY 1173 GCCGACTTGTGACATGACAGAGCTGGAGATCTTCTTCAACAACATTCCTACCGCAGC 1232
Db |||||
QY 141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160
Db |||||
QY 1233 TCTCAGCCCAACGGCTCTTCCGCTCATCCGTAGAGACGCGAGACCCCTCATCGAG 1292
Db |||||
QY 161 SerSerAlaAsnArgLeuPheArgValIleArgGluHisGlyAspProLeuIleGlu 180
Db |||||
QY 1293 GAGCTCAATCCCGGACCGCTTGGAGCTTGAGCGCTGAGGCGCGGCGAGGGGTGTGTGACC 1352
Db |||||
QY 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200
Db |||||
QY 1353 GACTTCGACGAGACGGGATGTGGACCTCATCTTGTCCATGGAGAGTCCATGGCTCAG 1412
Db |||||
QY 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220
Db |||||
QY 1413 CGCTGTCTGCTTCCGGGGCAATCAGGCTTCAACAACATGCTGGTGGAGTGGGCCA 1472
Db |||||
QY 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240
Db |||||
QY 1473 CGCACCGCTTTGGGACCTTTGCGAGGAGCTTAAGTGTCTGTCTACACCAAGAGAGT 1532
Db |||||
QY 241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260
Db |||||
QY 1533 GGGGCCCATCTGAGGATCATCGACGGGGCTCAGGCTACTGTGTGAGTGGAGCCCGTG 1592
Db |||||
QY 261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280
Db |||||
QY 1593 GCACACTTTGGCTGGGAGGATGAAGCAGAGTGTGGAGTGTGGCGCCAGATGGC 1652
Db |||||
QY 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerValGluValThrTrpProAspGly 300
Db |||||
QY 1653 AAGATGGTGGCGGAACTGGCGAGCGGGAGATGAATCACTGCTGGAGATCTCTAC 1712
Db |||||
QY 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320
Db |||||
QY 1713 CCGCGGATGAGACACACTTGGAGCCCGCCACTGAGTGTGGCCAGGATCTCC 1772
Db |||||
QY 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu----- 334
Db |||||
QY 1773 CAGCAGGAAATGGCCATTGTCATGGACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
Db |||||
QY 335 -----ThrProMetAsnAlaSerSerHisSerCysAl 346
Db |||||
QY 1833 CCTCGACACAGCCCGTATGTGTCAACACTATGGAAGCTTACAGTGGCGGACCAACAAG 1892
Db |||||
QY 346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366
Db |||||
QY 1893 AAGTGAAGTGGGGCTACGAGCCCAAGGATGGCACAGCTCGCTGGCTGGTGGAGC 1952
Db |||||
QY 366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly----- 383
Db |||||

QY 1953 CCTGTGTTGAAGATAGTAGACACCAACAGTGTGGGAAGAGCTTGTGCTGAATCACTGAA 2012
Db |||||
QY 384 -----LeuSerAlaSe 387
Db |||||
QY 2013 TCACCTGCCTTGAATCACCGCTGTGGAATACCTGTT 2046
Db |||||
QY 387 rHisArgAlaProAlaProProProProLeu 398
Db |||||

RESULT 10
US-10-188-832-56
; Sequence 56, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; APPLICANT: Bos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002330US
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-832-56

Alignment Scores:
Pred. No.: 2, 1e-121 Length: 418
Score: 1967.50 Matches: 385
Percent Similarity: 89.81% Conservative: 3
Best Local Similarity: 89.12% Mismatches: 10
Query Match: 48.51% Indels: 34
DB: 16 Gaps: 2

US-09-914-958b-35 (1-2177) x US-10-188-832-56 (1-418)
QY 753 ATGGACCTTGAGCCAGCTGACCTCTCCGGGGCATTCGCGCTCAGAGATGTGGTGCT 812
Db |||||
QY 1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaA 20
Db |||||
QY 813 GAGGCTGGGCTCAGCAATATATACAGGGGCCGAGCGTGTGGTGGCGCCATCTCAGC 872
Db |||||
QY 21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40
Db |||||
QY 873 AGCAGTGGCTCGATATCTTTCGACAAATGAGATGGCAAAAGTGGACATCTCTAT 932
Db |||||
QY 41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60
Db |||||
QY 933 CGGGCGATGGACCTTTGTGACGCTGGCGCAGTGTGTGGTGGACACCCACCCACAG 992
Db |||||
QY 61 ArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspProHisGln 80
Db |||||
QY 993 CATGGCGAGGTGTGCGCTGCTGATTCACCGGTGATGGCAAAAGTGGACATCTCTAT 1052
Db |||||
QY 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100
Db |||||
QY 1053 GGCACATGGAGTGGCCCGCCACCGCTCTATCTGCAATGAGCCATGGCAAGTGGAGTCCGC 1112
Db |||||
QY 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120
Db |||||
```



```
QY 1113 TTCCGGAGATCGCTCACCAAGTCTTCCATGATCCCTCCCTGTCGACGGTCAATCACC 1172
Db 121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140
QY 1173 GCGACATTTGACATGACAGGAGCTGAGATCTTCTCAACACATGCTGCTACCGGAGC 1332
Db 141 AlaAspPheAsnAspGlnGluLeuGluPhePheAsnAsnIleAlaTyrArgSer 160
QY 1233 TCCTCAGCCAAACCGCTCTTCCGCGTCATCCGTAGACAGACGAGACCCCTCATCGAG 1292
Db 161 SerSerAlaAsnArgLeuPheArgValIleArgGluHisGlyAspProLeuIleGlu 180
QY 1293 GAGTCAATCCCGCGAGCGCTTGAGAGCTGAGGCGCGGCGACAGGGGTGTGTGAC 1352
Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyThrGlyGlyValValThr 200
QY 1353 GACTTCACGAGAGCGGATCGCTGGACCTCATCTGTGCCATGAGAGTCCATGGCTCAG 1412
Db 201 AspPheAspGlyAspGlyMetLeuAspLeuLeuSerHisGlyGluSerMetAlaGln 220
QY 1413 CCGCTGTCCTCTTCCGCGGCAATCAGGGCTTCAACCAACTGGCTGCGAGTGTGTC 1472
Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240
QY 1473 CGACCCGGTTTGGGGCTTTCAGAGGAGCTTAAGTGTGTGCTCTACACCAAGAGT 1532
Db 241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260
QY 1533 GGGCGCCACCTGAGGATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTG 1592
Db 261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280
QY 1593 GCACACTTTGGCTGGGAAGGATGAAGCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
Db 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300
QY 1653 AAGATGTGAGCGGAACTGGTGGCGAGCGGAGAGTAACTCAGTGTGGAGATCCTCTAC 1712
Db 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320
QY 1713 CCCCAGGATGAGACACACTTCAGGACCCAGCCCTGAGTGTGGTGGCCAGGATTCCTC 1772
Db 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu----- 334
QY 1773 CAGCAGAAATGGCCATTCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
Db 335 -----ThrProMetAsnAlaSerSerHisSerCysAl 346
QY 1833 CCTCGAGACAGCCCGTATGTCAACACTATGGAAGCTACAGGTGCGGACCAACAG 1892
Db 346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366
QY 1893 AAGTGCAGTCGGGGTACGAGCCCAACGAGGATGGACAGCCTGCGTGGCTGTGTGAGC 1952
Db 366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly----- 383
QY 1953 CCTGTGTGAAGTATAGTACACCAAGTTGGGAAGAGCCTTGTGCTCCTGAATCACTGAA 2012
Db 384 -----LeuSerAlaSe 387
QY 2013 TCACTGCTTGAATACCGGCTCGAATACCTGTT 2046
Db 387 rHisArgAlaProAlaProProProProLeu 398
```

## RESULT 11

```
US-09-892-877-247
; Sequence 247, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
```

```
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-892-877-247
```

## Alignment Scores:

```
Pred. No.: 4,77e-90 Length: 306
Score: 1492.00 Matches: 291
Percent Similarity: 96.38% Conservative: 2
Best Local Similarity: 95.72% Mismatches: 7
Query Match: 36.79% Indels: 7
DB: 10 Gaps: 2
```

US-09-914-958B-35 (1-2177) x US-09-892-877-247 (1-306)

```
QY 1031 TGGCAAAAGTGGACATCGTCTATGCAACTGGAATGGCCCCACCGCCTCTATCTGCAAT 1090
Db 1 TrpGlnSerGlyHisArgLeuTrpGlnLeuGluTrpProProProLeuSerAlaAsp 20
QY 1091 GAGACCCATGGGAAGTCCGCTTCGGGACATCGCC---TCACCCAAAGTTCTCCATGCC 1147
Db 21 GluHisProTrpGluGlyProLeuProGly-ThrSerProSerProLysPheSerMetPr 40
QY 1148 CTCCTCTGT-CCGACGGTCTATCACC-CCGACTTTGACATGACAGGAGCTGGAGATC 1205
Db 40 oSerProValProHisGlyHisArgProThrLeuThrMetThrArgSerTrpArgGIl 60
QY 1206 -TTCTTCAACAAATTCCTTACCGCAGCTCTCAGCCCAACCGCCTCTTCGCGTCATCCG 1264
Db 60 ePhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgValIleAr 80
QY 1265 TAGAGACACGAGAGACCCCTCATCGAGAGCTCAATCCCGGCGAACCCCTGGAGCTGA 1324
Db 80 gArgGluHisGlyAspProLeuIleGluLeuAsnProGlyAspAlaLeuGluProGl 100
QY 1325 GGGCCGGGACAGGGGTGTGTGACCGACTTCGACGAGAGAGGGATGCTGACCTCAT 1384
Db 100 uGlyArgGlyThrGlyGlyValThrAspPheAspGlyAspGlyMetLeuAspLeuIl 120
QY 1385 CTGTGCTCCATGGAGAGTCCATGGCTCAGCCGCTGTCTCGTCTCCGGGGCAATCAGGGCTT 1444
Db 120 eLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPh 140
QY 1445 CAACACAACTGCTGCGAGTGTGCACGACACCCGGTTTGGGGCTTTCCCGAGGGAGC 1504
Db 140 eAsnAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaargGlyAl 160
QY 1505 TAAGTCTGCTCTACACCAAGAGAGTGGGGCCACCTCAGGATCATCCAGGGGGCTC 1564
Db 160 aLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySe 180
QY 1565 AGGTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGAAGAGATGAAGCCAG 1624
Db 180 rGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyLysAspGluAlaSe 200
QY 1625 CAGTGTGAGGAGTGAAGTGGCCAGATGGCAAGATGTGTGAGCCGGAACGTGCCACGGGGA 1684
Db 200 rSerValGluValThrTrpProAspGlyLysMetValSerArgAsnValAlaSerGlyGl 220
QY 1685 GATGAACCTCAGTCTGGAGATCTCTACCCCGGGGATGAGGACACACTTCAGGACCCAGC 1744
Db 220 uMetAsnSerValLeuGluIleLeuTyrProArgAspGluAspThrLeuGlnAspProAl 240
```



## RESULT 13

US-10-084-846A-6  
 ; Sequence 6, Application US/10084846A  
 ; Publication No. US20040006026A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEITNAUER, GABRIELE  
 ; APPLICANT: MUHLENKOPF, AGNES  
 ; APPLICANT: TREPZER, AXEL  
 ; APPLICANT: BECHTHOLD, ANDREAS  
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
 ; FILE REFERENCE: 1974-005  
 ; CURRENT APPLICATION NUMBER: US/10/084,846A  
 ; CURRENT FILING DATE: 2003-02-25  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
 ; PRIOR FILING DATE: 2001-02-25  
 ; NUMBER OF SEQ ID NOS: 120  
 ; SOFTWARE: PatentIn Ver. 3.2  
 ; SEQ ID NO 6  
 ; LENGTH: 19662  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces viridochromogenes  
 ; FEATURE:  
 ; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2.  
 ; OTHER INFORMATION: Start codon: 99a, Start position: nucleotide 1.  
 US-10-084-846A-6

## Alignment Scores:

Pred. No.: 3,83e-12 Length: 19662  
 Score: 318.00 Matches: 189  
 Percent Similarity: 32.75% Conservative: 36  
 Best Local Similarity: 27.51% Mismatches: 232  
 Query Match: 7.69% Indels: 230  
 DB: 15 Gaps: 44

US-09-914-958B-35 (1-2177) x US-10-084-846A-6 (1-19662)

Db	734	GlyArgSerAlaGlyProGlyGlyArgCysThrCysGlyArgCysGlyProLeuArg	753
QY	1332	CCCGCCTCAGGCTCCA---AGCGTCGCGGGATTGAGCTCTCTGATGAGGGGCTCTC	1276
Db	754	ArgGlyProAlaAlaCysArgArgArgProSerPro	768
QY	1275	CGTGCTCTTACGGATGACGCGAAGAGCGGTTGG---CTGAGGAGTGGGTAGGCAA	1219
Db	769	---GlyThrArgProTrpSerCysArgSerCysSerHisArg	781
QY	1218	TGTTGTTGAAGAAGATCTCAGCTCTCGTCAATTGTCAAGTCGGCGGTGAGACCGTGC	1159
Db	782	CysSerArgSerAlaGlyProGlyGlyArgAlaAlaGlyArg	798
QY	1158	GGACAGGG---AGGCGATGG---AGAACTTGGTGGCGCATGT---1120	
Db	799	GlyAlaGlyTrpAlaSerProCysArgGlyTrpArgArgArgTrpArgThrSer	818
QY	1119	---CCCGAAGCGGACCTTCCCATGGTGCTCATTTGCAGATAGAGGGCGGT	1072
Db	819	AlaProTrpArgProGlyThrGlyTrpProArgGlyTrpArg	835
QY	1071	---GGGGCCATTCAGTTGCCATAGACGATGCCACTT	1036
Db	836	AlaArgArgProProArgAlaGlySerProSerThrTyrArgArgTrpCysTrpSer	855
QY	1035	TGCCATCACGGTTGAAGTCAGCCA---GGCGCACACCTC	1000
Db	856	ProSerArgLeuSerSerArgProArgTrpThrAsnArgGlyArgGlyArgGlyArg	875
QY	999	GCCCATGCTGGTGGGGTCTGCCA---CACACGACCTGGCGG	961
Db	876	SerArgSerTrpSerArgProTrpArgProSerTrpSerArgArgProTrpPro	895
QY	961	---GGCGCACACCTC	961
Db	896	ArgArgValArgValThrSerTrpArgTrpSerThrArgCysArgProProArgArg	915
QY	960	---CAGCGTCCACAAAGGTGCCATCGCCCGTTGTGCGAAAGAAAGTAGGCCCAT	907
Db	916	ArgArgArgProTrpArgCysValArgProCysArgGlyPro	930
QY	906	TCTCATTTGCGAAGATATCCAGGCACTGCTGCTGAGGATGG---GGCCACCGCTGA	850
Db	931	---CysArgArgArgArgGlyThrCys---TrpArgArgProAlaArg	945
QY	849	CGCTCGGCCCTGTATTGCTACCCAGCCCTCAGCAGCCATCTCTGAGGCGCA	790
Db	946	ArgSerGlyArgProSerArgCysArgArgProGlyArg	959
QY	789	GAATGCCCGGAGAGGTCACTGGCCTCAGGGTCCATTTCATAGAGGCATCAGGCGCCA	730
Db	960	TrpCysProGlyArgSerCysSerAlaAlaGlyArgLeuArg	976
QY	729	CATTACCTAGCGTAAATTTGGCAATGATAGTAGAGTAGCGTCCAGAGCCCTTCTGTCCA	670
Db	977	SerTrpArgArgArgArg---ArgCysArg---ArgPro	987
QY	669	CACAGGCCACAGAGCGTCCGGCAAGAGGCTGGCCACACCGGCGCACGTTGACCTCAT	610
Db	988	CysArgProArgAlaCysArgGlyArgGlyArgPro	999
QY	609	CGCTCAGGATGTTCCCGCCGTTATTGCGGAACCTTGAACAACCTGTCGTTGACCTGG	550
Db	1000	---GlyArgCys---CysArgCysArgTrp	1007
QY	549	CCACCCCGGAGAGCAATTATTGTTGTGAGGAAGTAGATCTCTCCCGCGCTCCCGGT	490
Db	1008	ArgArgSerArgArgPro-----GlyArgArgSerArgCysGlyArg	1024
QY	489	CGATGTCGAGGCTGTGACCCCAATGGCGTTCCCTGCGCGTCCCGCAGCGGTAGTAGG	430
Db	1025	ArgCysArgArgThr---ProSer-----LeuProAlaArgProSerProGlyArgCys	1041



```
QY 817 CTGGGTCAGCAATATACAGGGGCGGAGCGTACGGTGGCCCGCCATCC----- 867
Db 6759 -----SerTrpArgProAlaProArgAla 6766
QY 868 ---TCAGCAGCAGTGCCTCGGATA----- 888
Db 6767 LysSerAlaAlaProSerArgTrpArgProThrArgCysArgValSerGlyArgSer 6786
QY 889 ---TCTTCTCGCAATAGAGATGGGCTAACTTCCT----- 922
Db 6787 ArgSerAlaAlaArg-TipCysGlyTrpProProGlyArgSerArgTrpProPr 6806
QY 923 -----TTTCCACACCGGGCGA----- 940
Db 6806 oThrArgTrpProGlySerSerMetProArgProGlyArgProArgTrpSerPr 6826
QY 941 ---TGGCACCTTTGTGACGCTCGGCCAGTGTGG-----TGTGG 978
Db 6826 ovalTrp-----ArgArgGlyArgCysTrpThrArgAlaSerThrGlySerAs 6842
QY 979 ACGACCCCGCCACCATGGCGAGGTGCGCCCTGGCTGACTTCAACCGTGATGGCAAG 1038
Db 6842 nArgProThrProAlaTrpSer-----ProGly----- 6851
QY 1039 TGGACATCGTATGGCACTGGGAATGGCCCGCCCGCTCTATCTGCAATGAGCACCC 1098
Db 6852 -----GlyTrpPro-----Ar 6855
QY 1099 ATGGGAAGTCCCGTTCGGGACATCGCTCACCCAAAGTTCTCCATGCCCTCCCTGTC 1158
Db 6855 gTrpAlaTyProArgPro--valSerIleGlyAlaSerSerAspCysProProValSer 6874
QY 1159 GCA-----CGGTATCATCCCGCGACT---TTGACAAATGACCGAGCTGGAGA 1203
Db 6875 ValTyrCysThrLeuArgCysSerArgProThrProArgProGlyProArgCysTrpThr 6894
QY 1204 TCTTCTTCAACAATGGCTACCGAGCTCCTCAGCCACCGCTCTTCGGCTGCTATCC 1263
Db 6895 ProCysArgSerArgProProSer-----TrpAlaSerArgSerSerAlaTrpSer 6912
QY 1264 GTAGAGACGACGAGACCCCTCA-----TCGAGGAGC 1296
Db 6913 CysThrSerThrArgSerProSerThrAlaArgHisArgArgArgArgSerArgSer 6932
QY 1297 ---TCAATCCGCGG-----ACGCTTTGAGCCTTGAGGCGCGGCGGCACAG 1338
Db 6933 ArgSerIleProAlaSerProThrProCysThrProTrpSer---ArgThrGlyArgAsp 6951
QY 1339 GGGGTGTGTGACCGACTTCGACGGACAGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
Db 6952 AlaArgTrpArgAlaCysAlaAlaCysAlaAlaThrArgTrpSerSerSerArgProArg 6971
QY 1399 AGTCCATGGCTACGCCGCTGCTGCTCCGGGCAATCAGGGCTTCAACACAACTGGC 1458
Db 6972 ThrPro-----ThrSerArgAlaAlaThrArgThrArg 6982
QY 1459 TGGCAGTGTGCCACCGCA-----CCCGTTTGGGCGCT 1491
Db 6983 ---ThrTrpCysArgSerArgValCysArgArgSerCysValProGlyCysSerThr 7001
QY 1492 -----TTGCCAGGGAGCTAAGTCTGCTACACAGAGAGAGTGGGGCCC---ACC 1542
Db 7002 ArgCysAlaSerArgSerArgArgCysAspSerProProArg-----ProCysThr 7019
QY 1543 TGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGG-----AGCCCGTGG 1593
Db 7020 PheAlaAlaArgTrpTrpSerArgGlySerThrArgArgTrpTrpSerAlaAlaGlyTrp 7039
QY 1594 CACACTTTGCTGGGGAAGATGAAGCCAGCAGTGTGGAGGT----- 1636
Db 7040 ArgSerAlaSerAlaGlyThrCysArgProThr--SerSerGlySerCysProProSerAl 7059
```

```
QY 1637 -----GACGTGGCCAGATGGCAAGATGTTGACCC 1665
Db 7059 aThrSerSerIleThrProAsnAlaSerArgAsnThrArgArgTrpAlaMetArgPr 7079
QY 1666 GGA-----CGTGGCCAGCGGGGAGATGAACTCAGTGC 1698
Db 7079 oGluArgThrAspSerValThrHisValSerArgGlyArgGlyGluArgAlaAlaPr 7099
QY 1699 TGGAGATCCTCTACCCCGGATGAGGACACACTTCAGGACCCAGC----- 1744
Db 7099 oAlaGlyPro-----GluThrGlyLysProGluLysHisAlaSerTh 7113
QY 1745 -----CCCACTGGAGTG----- 1756
Db 7113 rIleValValGlyArgArgHisSerMetProSerGlyValProLeuLysArgProAspLe 7133
QY 1757 -----TGG----- 1759
Db 7133 uThrAlaLeuThrAlaArgCysGlyIleArgArgSerTyrTrpTrpAlaGlyLeuAlaAr 7153
QY 1760 -----CCAAGGATTCCTCCAGCAGGAAATGGCCATTCATGG 1797
Db 7153 gGlyCysGlyArgLeuArgIleProArgProAsnAsnTrpSerArgTrpProIleGluLe 7173
QY 1798 ACACCAATGAATGCATCCAGTTCCTCCATTCGTGTGCC----- 1834
Db 7173 uPheSerProMetTyrTrpIleProSerArgProGluLysArgProGluLysArgSerGlyLeuSe 7193
QY 1835 ---TCGAGACACGCCGCTATGTGTCAACACCTATGAGCTACAGGTGCGCGGACCAACA 1890
Db 7193 rSerGlyArgProArgArgSerArgLeuProTrpGluThrAlaArgSerSerAlaTr 7213
QY 1891 AGAAGTCAGTCGGGGCTACGAGCCCAACGA---GGATGGCACAGCTCGCTGGGCTGGT 1947
Db 7213 pThrSerProThrSerSerArgThrProArgAlaGlyTrpProThr----- 7228
QY 1948 GAGGCCCTGT-----GTTGAAGATAGTGACACCAAGTTGGGAAGACCTTGGTC 1998
Db 7229 ---ProCysSerCysProValThrSerSerAlaThrThrThrSerSerCysThrTrpAl 7247
QY 1999 CCTGAATCACTGAATCACTGCTGATCAATCAGCCCTGGAATACC 2042
Db 7247 aThrThrSerTrpThrAlaSerSer-ThrSerTrpArgThr 7261
```

## RESULT 15

```
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILING DATE: 2001-08-24
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PR
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
```

```
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5
```

Alignment Scores:

Pred. No.: 2,32e-10 Length: 19723  
 Score: 291.00 Matches: 220  
 Percent Similarity: 30.56% Conservatives: 48  
 Best Local Similarity: 25.09% Mismatches: 297  
 Query Match: 7.04% Indels: 313  
 DB: 15 Gaps: 48

US-09-914-958B-35 (1-2177) x US-10-084-846A-5 (1-19723)

QY 1953 GGCCTCACAGCCAGCC---AGGCTGTGCTTCCTCTGGCTGGCTGGTACGCCGCTGCTG 1897  
 Db 6645 GlyAlaThrSerProArgProGlyProSerAlaProProTriAla---ProProTriAla 6663  
 QY 1896 ACTCTCTGTGGTCCGGC-----1879  
 Db 6664 ThrAlaAlaTriPThrGlyArgSerSerArgThrGlyThrSerLysSerCysArgPro 6683  
 QY 1879 -----1879  
 Db 6684 ThrProGlySerThrAlaArgCysGlyValProSerProGlyArgGluArgProSerAla 6703  
 QY 1878 -----ACCTGTAGCTTCATAGTGTTCACATACAGGCTGTCTCGAGGCGCACAGA 1825  
 Db 6704 AlaThrThrSerThrProThrArgCysTriAlaArgProAlaAlaSerArgAlaAlaThr 6723  
 QY 1824 ATGGGAACGGATGCTCATTGCTGCT-----1798  
 Db 6724 CysGlyCysGlyThrAlaArgTriPThrThrAlaAlaArgThrArgAlaAlaThrPro 6743  
 QY 1797 -----CCATGC-----AATGGCCATTTTCCTGCT 1774  
 Db 6744 SerThrAlaTriPArgSerSerArgProProCysTriPArgSerArgSerTriPArgProAla 6763  
 QY 1773 GGGAGAACTCTGGCCACACTCCAGTGGGCTGGTCTGAGTGTCTCATCCCGGG 1714  
 Db 6764 ProArgAlaLysSerAlaAlaProSerArgThrArgProThrArgCysArgValSer 6783  
 QY 1713 GGTAGAGGATCTCCAGCACTGCTCATCTCCCGCTGG-----1675  
 Db 6784 GlyArgSerArgSer-----SerAlaArgAlaArgTriPArgCysGlyTriPProProGlyArg 6801  
 QY 1674 -----1674  
 Db 6802 SerArgThrProProThrArgThrProProGlySerSerMetProArgProGlyArgPro 6821  
 QY 1656 TCTTGCATCTGCCACCTCACCT-----CCACACTGGCTGGCTTCATCTCTCC 1609  
 Db 6822 Arg-----ArgThrSerProValTriPArgArgGlyArgCysTriPThrArgAlaSer 6838  
 QY 1608 CCAGGGCCAAAGTGTGCCAGGCTCCATCTCACACAGGTAGC-----CTGAGC 1561  
 Db 6839 ThrGlySerAsnArgProThrProAlaTriPThrSerProGlyGlyTriPProArgTriPAlaTriP 6858  
 QY 1560 CCCCGTCGATGATCCAGTGGG-----CCCACTCTCTCTGGTGTAGA 1516  
 Db 6859 ProArgProValSerIleGlyAlaSerSerAspCysProProValSerValTriPThrCysThr 6878  
 QY 1515 GCACGACCTTAGCTCCCTGGGAAAGCCCAACCGGTCGCTGGCACCCTCGAGCC 1456  
 Db 6879 LeuArgCysSerArgProThrProArgProGlyProArgCysTriPThrPro-----6895  
 QY 1455 AGTTGTGTGTGAAGCTGATTCCTCCCGGAAGACGACAGGCTGAGCCATGCACTCTC 1396  
 Db 6896 -----CysArgSerArgProProSerTriPAlaSerArgSerAlaTriPThrCys 6913  
 QY 1395 CATGGGACAGATGAGTCCAGCATCC-----CGTCTCGTCTGAATCGG---1351  
 Db 6914 ThrSerThrArg---SerProSerThrAlaArgHisArgArgArgSerArgSer 6932  
 QY 1350 -----TCACACACCCCTGTGCCCCG-----1329  
 Db 6933 ArgSerIleProAlaSerPro-ThrProCysThrProTriPThrSerArgThrGlyArgAspAl 6952

QY 1328 -----GCCCTCAGGCTCCAAAGCGCTGCCGGAGTTGAGCTCTCGATGAGGGGT 1279  
 Db 6952 aArgTriPArgAla-CysAlaAlaCysAlaAlaTriPThrSerSerArgProArgArgT 6972  
 QY 1278 CTCCTGTCTCTACGATGACGCGAGAGAGCGGTTGGCTGAGAGCTCGGTAGGCA 1219  
 Db 6972 hrProThrSerArgAlaAlaAlaThrArgThr-TriPArgSer---Arg 6988  
 QY 1218 TCTTGTGAAGAGATCTCCAGCTCTGCT-----CATTTCAAAGTCGGCGTGA 1168  
 Db 6989 ValCysArgArgSerCysValProGlyCysSerThrArgCysAlaSerArgSer---7007  
 QY 1167 TGACCGTGGCAGCAGGGAGGCGATGGAGAACTTGGGTGAGCGATGTCGCCGAAACGGA 1108  
 Db 7008 -----ArgArgArgCys-----AspSer 7013  
 QY 1107 CTTTCCATGGGTGCTCATTTGCGATAG-----AGGCGGTGG 1070  
 Db 7014 ProProArgProCysThrPheAlaAlaAlaArgTriPThrSerArgGlySerThrArgArgT 7033  
 QY 1069 GGGCCATTCAGTTGCCATAGCAGATGCTCCACTTGGCCATCAGGTTGAAGTCAGCCNG 1010  
 Db 7034 -----TriPAlaAlaGlyTriPArgSerAlaAla---7044  
 QY 1009 GCGACACTCGCCCATGCTGGTGGGTCTCTCCACACAGCACTGCGCGAGCGTCCACA 950  
 Db 7045 GlyThrCysArgProThrSerSerGlySerCysProProSerAlaAlaThrSerSerIleThr 7064  
 QY 949 AAGTGCCTATCGCCCGGTTGTGAAAGAGAGTTAGGCCCATCTCATTTGTCGAGAG 890  
 Db 7065 ProAsnAlaSerArgAsnThrArgArgArgThrAlaMetArgProGluArgThrAspSer 7084  
 QY 889 ATATCCGAGGCACTGCTGCTGAGGATGGGCGCCACGCTCAGCGCTCGGCCCTC-----836  
 Db 7085 ValThrHisValSerArgGlyArgGlyGluArgAlaAlaProAlaGlyProGluThr 7104  
 QY 836 -----836  
 Db 7105 GlyLysProGluLysHisAlaSerThrIleValGlyArgArgHisSerMetProSer 7124  
 QY 835 -----GTTATTTGTTGACCCAGCCTCAGCAGCCACATCTCTGAGCCGCAATGC-----784  
 Db 7125 GlyValProLeuLysArgPro-----AspLeuThrAlaLeuThrAlaArg-CysGlyI 7142  
 QY 783 -----CCCGG 779  
 Db 7142 eArgArgSerTy-TriPAlaGlyLeuAlaArgGlyCysGlyArgLeuArgIleProAr 7162  
 QY 778 GAGAGTCACTGCTCCTCAGGCTCCATTCAA-----TGAGGGCA 740  
 Db 7162 gProAsnAsnTriPThrSerArgTriPProIleGluLeuPheSerProMetTyTriPProSe 7182  
 QY 739 TCAGGGCCACATTCAGGTAGCGTAAT-----712  
 Db 7182 rArgArgProGluTyArgArgSerGlyLeuSerSerGlyArgProArgArgSerAr 7202  
 QY 711 -----TGCAAA---TGATAGTAGAGTAGCTCCAGAGCCCTTCTGTCCACACAGGCC 662  
 Db 7202 gLeuProTriPAlaArgSerSerAlaTriPThrSerProThrSerSerArgThrPr 7222  
 QY 661 ACAGAGCTCGGCAAGAGCGTGGCCACACACGCGGCCAGCTTGACCTCATCGTCAGG 602  
 Db 7222 oArgAla-----GlyTriPPro-----ThrPr 7229  
 QY 601 ATGT-----CTTCCACCGGTTATTGCGAACTTGAACAACCTTGTGGTGTACGTGG-----550  
 Db 7229 oCysSerCysProValThrSerSerAlaThr---ThrThrSerSerCysThrTriPAlaTh 7248  
 QY 549 -----CCACCCC 542  
 Db 7248 rThrThrSerTriPAlaSerSerThrSerTriPArgThrSerAlaAlaThrSerProPr 7268  
 QY 541 GAGAAGGCATTATTGTGTGTGAGGAAGTAGATCTCTCCCGCGGCTCCCGTCGATCG 482

```
Db 7268 oArgArg-----SerCysSerProGlySerProThrArgAlaG1 7281
QY 481 CAGGCTGTGACC-----CAATGGCGTTCCCTGCCGG 449
Db 7281 ySerAlaTTrpProAsnLeuThrArgThrAlaGlyProArgSerTrpArgSerProArgTh 7301
QY 448 TCCGGCAGCGGTAGTAGGTGAGTGC-----GCTCATCGACCG-----CGATG 404
Db 7301 rProGlyAlaThrTrpArgSerSerAlaCysThrProSerAlaArgProCysThrArgPr 7321
QY 403 TTCACACGCGCTTCTGGCCCGTCTACTTCAGAACACAGGTTGGTCCATTGTACCCC 344
Db 7321 oTrpProAlaSerGlyArgProGlyGlyThrSer-----TrpArgSerProThrPr 7338
QY 343 GCCACGACGATCTCAAAGTCCCATCATGTGTCACATCAGTAACTGCCACACCATAGTTG 284
Db 7338 o-----CysSerGlyTrpProThrThrGlyThrArgSer--AlaProArgSerThrG1 7355
QY 283 AGCTGGGTGGGATTACTGTCTAGTCAGGAGGAGAACTAGTTGGTGTGACTGCAGTGAAC 224
Db 7355 yThrGlyArgThr-----SerArgAlaSerProThrSerTrpArgThrGlyThrCy 7372
QY 223 ATGGGTT-----CAGCCCGCTGGACCCCTCAGTGATGGGACAGAACACAGACGACGAGC 170
Db 7372 sTrpSerProTrpArgProAlaSerThrAlaArg---Trp-----ThrArgProAlaAs 7389
QY 169 AGAACCGGTAACTCTGGACATGCCGGGTCTAGCGCTCGGAGCCATCTCCCGCTCTCG 110
Db 7389 nTrpTrpAlaGluSerTrpTrpArgArgGlyArgSerProAlaProAlaSerSerVa 7409
QY 109 GCCCGCGCGCTAGGGCGGTGGGAAGCGGGCGCTCGCTGCCGCTCTGCCGCGCGCGCG 50
Db 7409 lPro-----ProSerSerAlaArgThrArgTrpCysArgThrProth 7423
QY 49 CCGCTGTCTGTCTCCACAGCCCGCTCCCGGGCTGGCTCGAGC 7
Db 7423 rSerVal-----ProSerPro---ProSerArgAlaThrAla 7434
```

Search completed: August 3, 2004, 20:08:23  
Job time : 221 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 12:54:35 ; Search time 8286 Seconds  
(without alignments)  
11387.616 Million cell updates/sec

Title: US-09-914-958B-35

Perfect score: 2177

Sequence: 1 cggagagctcgagccagcc.....tcctcttgaaaaaaaaaa 2177

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2177	100.0	2177	6	BD269644	BD269644 Human sec
2	2043.8	93.9	2089	9	AX357116	AX357116 Sequence
3	1940.4	89.1	2681	9	HS276171	AJ276171 Homo sapi
4	1934	88.8	2889	9	AK057190	AK057190 Homo sapi
5	1913.4	87.9	2413	6	AX427468	AX427468 Sequence
6	1893	87.0	2263	6	AX202100	AX202100 Sequence
7	1808.8	83.1	2589	6	AX357100	AX357100 Sequence
8	1807.8	83.0	2039	6	AX357115	AX357115 Sequence
9	1807.8	83.0	2145	6	AX357117	AX357117 Sequence
10	1794.4	82.4	1959	6	AX470036	AX470036 Sequence
11	1794.4	82.4	2507	6	AX470035	AX470035 Sequence
12	1794.4	82.4	2507	9	HS279016	AJ279016 Homo sapi
13	1793.4	82.4	1947	6	AX470056	AX470056 Sequence
14	1793.4	82.4	1947	9	HS279016	AJ279016 Homo sapi
15	1476.4	67.8	1973	10	MMU421516	AJ421516 Mus muscu
16	1438	66.1	2147	9	BC034245	BC034245 Homo sapi
17	1408.8	64.7	2178	6	AX876177	AX876177 Sequence
18	1408.8	64.7	2178	6	BD155992	BD155992 Primer fo
19	1408.8	64.7	2178	9	AK001182	AK001182 Homo sapi
20	1345.8	61.8	1501	6	AR339541	AR339541 Sequence
21	1137.8	52.3	1558	10	BC024472	BC024472 Mus muscu
22	1025.6	47.1	1143	6	BD205640	BD205640 97 human
23	788	36.2	789	6	AR379939	AR379939 Sequence
24	556.8	25.6	608	6	AX867169	AX867169 Sequence
25	556.8	25.6	608	6	BD147231	BD147231 Primer fo
26	449	20.6	700	10	RNU78304	U78304 Rattus norv
27	229	10.5	492	11	G27517	G27517 human STS S
28	225	10.3	146124	9	AL358938	AL358938 Human DNA
29	207	9.5	161235	9	AL139239	AL139239 Human DNA
30	207	9.5	172383	2	AC015630	AC015630 Homo sapi
31	174	8.0	300750	1	AP006576	AP006576 Gloeobact
32	164	7.5	163348	10	AC119236	AC119236 Mus muscu
33	164	7.5	243770	2	AC120763	AC120763 Rattus no
34	164	7.5	252404	2	AC097117	AC097117 Rattus no
35	161.6	7.4	395	11	G25373	G25373 human STS E
36	161.2	7.4	172383	2	AC015630	AC015630 Homo sapi
37	158	7.3	164766	10	AC144792	AC144792 Mus muscu
38	148.2	6.8	409	11	HSU123C5	AL159715 STS from
39	145.4	6.7	210385	10	AC106128	AC106128 Rattus no
40	142.2	6.5	189196	10	AL603804	AL603804 Mouse DNA
41	130.2	6.0	229480	5	AL929150	AL929150 Zebrafish
42	114.8	5.3	110000	2	AC099306	Continuation (6 of
43	110.4	5.1	743	11	BV022941	BV022941 S212P6878
44	109	5.0	707	6	AR379664	AR379664 Sequence
45	101.4	4.7	166810	2	BX005216	BX005216 Danio rer

# ALIGNMENTS

RESULT 1	BD269644	2177 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD269644				
DEFINITION	Human secretory proteins.				
ACCESSION	BD269644				
VERSION	BD269644.1	GI:33079412			
KEYWORDS	JP 2002537805-A/13				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 2177)				
AUTHORS	Tang,T.Y., Lal,P., Baughn,M.R., Yue,H., Young,J.A., Lu,D.A.M. and Azimzai,Y.				
TITLE	Human secretory proteins				







```

CDS
183..2135
/ gene="ASPICI"
/ standard_name="acidic secreted protein in cartilage"
/ notes="putative alternative start, less identity to the
Kozak rule"
/ codon_start=1
/ evidence=experimental
/ product="ASPIC"
/ protein_id="CAB98269.1"
/ db_xref="GI:9368808"
/ db_xref="GOA:Q9NQ78"
/ db_xref="SPTREMBL:Q9NQ78"
/ translations="MLPFLILLFLPITEGQSAEPMFTATVNSVLPPDYDSNPQLN
YGVAVTDVHDGDFEIVVAGYNGENLVKIDRAQKRLVNIADVERSSPYVALDRQGN
AIGVTACDIDGDEEIIYFNTNNAFSGVAITYTDKQKFRNNRWEDILSDENVARGV
ASLPAGRSVACDKRGSRYSIYANYAYNGVPDALIENDPEASDILSRGILALRDVA
AEAGVSKYTGGRVSGVPISSASDIFCDNENGNPFLFHNRGDGFVDAASAGVDD
PHQHGVALANDNRDKVDIVYGNWNGPRLYLQWSTHGKPRFDIASPFMSRSPV
RTVITADFNDQDELEIFNNIAVRSSANLFRVIREHGDPLIEELNPGDALEPRGR
GTGAVTDFPDGCMILLILSHGSRMAQPLSVFRGNQGFNNWLRVPRTRFGAPFAGA
KVLYTKSAHLRIIDGGGYLCMEPEPVAHFGLGKDEASSVEVTPWPDGRMVRNVAS
GENWSVLEILYLPDEDTLODPAPLECGQSFQOENGHCMDNECIQFPFVCPDRDKPVC
VNTYGSYRCRTNKKCRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAAGA
ATAAPVLVDGDLNLGSLVVKESCEPSC"
183..230
sig_peptide
/ gene="ASPICI"
/ notes="putative alternative start, less identity to the
Kozak rule"
231..2132
mat_peptide
/ gene="ASPIC"
/ product="ASPIC"
/ standard_name="acidic secreted protein in cartilage"
2136..2681
3'UTR
/ gene="ASPICI"
2642..2647
polyA_signal
/ gene="ASPICI"
ORIGIN
Query Match 89.1%; Score 1940.4; DB 9; Length 2681;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGGGAGGCTCGAGGCGCGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCGGG 60
DB 28 CGGGAGGCTCGAGGCGCGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGGG 87
QY 61 GCAGAGGCGGACGAGGCGCGGCTTCCACGCCCTTAGCGGCGGCGGCGAGCGGGA 120
DB 88 GCAGAGGCGGACGAGGCGCGGCTTCCACGCCCTTAGCGGCGGCGGCGAGCGGGA 147
QY 121 GGATGGCTCGGAGCGGTGACCCGGCATGTCAGGATGTTACGGTTCCTGCTGCTCT 180
DB 148 GGATGGCTCGGAGCGGTGACCCGGCATGTCAGGATGTTACCGTTCCTGCTGCTCT 207
QY 181 GGTTCCTGCCCATCACTAGGCGTCCACGCGGCTGAAACCCATGTTCACTGCAGTCAACA 240
DB 208 GGTTCCTGCCCATCACTAGGCGTCCACGCGGCTGAAACCCATGTTCACTGCAGTCAACA 267
QY 241 ACTAGTTCGCTCTGCTACTATGACAGTAATCCACCCAGTCAACTATGTTGGCCAG 300
DB 268 ACTAGTTCGCTCTGCTACTATGACAGTAATCCACCCAGTCAACTATGTTGGCCAG 327
QY 301 TTACTGATGAGCAATGATGGGACTTTGAGNATCGTCTGGCGGGGTACAATGGACCCA 360
DB 328 TTACTGATGAGCAATGATGGGACTTTGAGNATCGTCTGGCGGGGTACAATGGACCCA 387
QY 361 ACTGTTTCTGAAGTATGACCGGCGGCGGAGCGGCTGGTGAACATCGCGTTCGATGAC 420
DB 388 ACTGTTTCTGAAGTATGACCGGCGGCGGAGCGGCTGGTGAACATCGCGTTCGATGAC 447
QY 421 GCAGTTCACCTTACTACGCGCTGCGGGACCGGCGGGAACGCCATTGGGGTTCACAGCCT 480
DB 448 GCAGTTCACCTTACTACGCGCTGCGGGACCGGCGGGAACGCCATTGGGGTTCACAGCCT 507

```

```

QY 481 GCGACATCGACGGGACGCGCGGAGGAGATCTACTTCTCAACACCAATTAATGCCTTCT 540
DB 508 GCGACATCGACGGGACGCGCGGAGGAGATCTACTTCTCAACACCAATTAATGCCTTCT 567
QY 541 CGGGGCTGGGCAACGTACACCGACAAGTTGTTCAAGTTCGCGCAATAACCGGTGGGAAGACA 600
DB 568 CGGGGCTGGGCAACGTACACCGACAAGTTGTTCAAGTTCGCGCAATAACCGGTGGGAAGACA 627
QY 601 TCCTGACCGCATGAGGTCAAAGTGGCCGTGTGTGGCCAGCCTCTTTGCCGGAAGCTCTG 660
DB 628 TCCTGACCGCATGAGGTCAAAGTGGCCGTGTGTGGCCAGCCTCTTTGCCGGAAGCTCTG 687
QY 661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTCGCAATTAACGCT 720
DB 688 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTCGCAATTAACGCT 747
QY 721 ACGGTAATGTGGCCCTGTATGCCCTCATTTGAATGGACCTTGAGGCGAGTGAACCTCTCC 780
DB 748 ACGGTAATGTGGCCCTGTATGCCCTCATTTGAATGGACCTTGAGGCGAGTGAACCTCTCC 807
QY 781 GGGGCATTCTGGCGCTCAGAGATGTGCTCTCAGACGAGTGCCTCGGATATCTTCTGCGACA 840
DB 808 GGGGCATTCTGGCGCTCAGAGATGTGCTCTCAGACGAGTGCCTCGGATATCTTCTGCGACA 867
QY 841 GCGAGGCGCTCAGCGTGGGCCCCCATCTCAGACGAGTGCCTCGGATATCTTCTGCGACA 900
DB 868 GCGAGGCGCTCAGCGTGGGCCCCCATCTCAGACGAGTGCCTCGGATATCTTCTGCGACA 927
QY 901 ATGAGAAATGGGCTTAATCTCTTTTTCACAAACCGGGCGATGGACCTTTGTGGAGCTG 960
DB 928 ATGAGAAATGGGCTTAATCTCTTTTTCACAAACCGGGCGATGGACCTTTGTGGAGCTG 987
QY 961 CGGCCAGTGTCTGTGTGGACGACCCCGACAGCATGGCGAGTGTGCGCTGCTGCTGACT 1020
DB 988 CGGCCAGTGTCTGTGTGGACGACCCCGACAGCATGGCGAGTGTGCGCTGCTGCTGACT 1047
QY 1021 TCAACCGTGTATGGCAAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCTCT 1080
DB 1048 TCAACCGTGTATGGCAAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCTCT 1107
QY 1081 ATCTGAAATGAGACACCAATGGGAAGTGTGCTTTCGGGACATCGGCTCAACCAAGTTCT 1140
DB 1108 ATCTGAAATGAGACACCAATGGGAAGTGTGCTTTCGGGACATCGGCTCAACCAAGTTCT 1167
QY 1141 CCATGCGCTCCCTGTCCGACGCTCATACCGCGCATTTTGACAAATGACAGAGCTGG 1200
DB 1168 CCATGCGCTCCCTGTCCGACGCTCATACCGCGCATTTTGACAAATGACAGAGCTGG 1227
QY 1201 AGATCTTCTTCAACAAATGGCTACCGAGCTCTCTCAGCCAAACCGCTCTTCCGGCTCA 1260
DB 1228 AGATCTTCTTCAACAAATGGCTACCGAGCTCTCTCAGCCAAACCGCTCTTCCGGCTCA 1287
QY 1261 TCGGTAGAGACACGAGACCCCTCATCGAGAGCTCAATCCCGGCGAGCGCTTGAGC 1320
DB 1288 TCGGTAGAGACACGAGACCCCTCATCGAGAGCTCAATCCCGGCGAGCGCTTGAGC 1347
QY 1321 CTGAGGCGCGGGGACAGGGGGTGTGTGACCGCTTCGACGAGACGGGATGCTCGGAC 1380
DB 1348 CTGAGGCGCGGGGACAGGGGGTGTGTGACCGCTTCGACGAGACGGGATGCTCGGAC 1407
QY 1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCGCTGTCCGCTTTCGGGGCAATCAG 1440
DB 1408 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCGCTGTCCGCTTTCGGGGCAATCAG 1467
QY 1441 GCTTCAACAACTGGCTGGAGTGTGCGACAGCACCGGTTTGGGGCTTTGGCAGGG 1500
DB 1468 GCTTCAACAACTGGCTGGAGTGTGCGACAGCACCGGTTTGGGGCTTTGGCAGGG 1527
QY 1501 GAGCTAAAGTGTGTCTTACACCAAGAGAGTGGGGCCACCTGAGGATCATCGAGGG 1560
DB 1528 GAGCTAAAGTGTGTCTTACACCAAGAGAGTGGGGCCACCTGAGGATCATCGAGGG 1587

```

```
Qy 1561 GCTCAGGCTACCTGTGTGAGATGAGGCGCCGTGGCAGACACTTTGGCTGGGAAGGATGAAG 1620
|
|
|
Db 1588 GCTCAGGCTACCTGTGTGAGATGAGGCGCCGTGGCAGACACTTTGGCTGGGAAGGATGAAG 1647
|
|
|
Qy 1621 CCAGCAGTGTGAGGTGACGTGGCCAGATGCAAGATGGTGAAGCGGGAACCTGGCCAGCG 1680
|
|
|
Db 1648 CCAGCAGTGTGAGGTGACGTGGCCAGATGCAAGATGGTGAAGCGGGAACCTGGCCAGCG 1707
|
|
|
Qy 1681 GGGAGATGAATCAGTGTGAGATGCTCTACCCCGGGATGAGGACACACTTCAGGACC 1740
|
|
|
Db 1708 GGGAGATGAATCAGTGTGAGATGCTCTACCCCGGGATGAGGACACACTTCAGGACC 1767
|
|
|
Qy 1741 CAGCCCCACTGGAGTGGCCAGAGATTCCTCCAGCAGGAAATGGCCATTCGATGGACA 1800
|
|
|
Db 1768 CAGCCCCACTGGAGTGGCCAGAGATTCCTCCAGCAGGAAATGGCCATTCGATGGACA 1827
|
|
|
Qy 1801 CCAATGAATGATCCAGTTCCTCCATTCGTCGATGAGGACCAAGCCCTATGTGTCAACA 1860
|
|
|
Db 1828 CCAATGAATGATCCAGTTCCTCCATTCGTCGATGAGGACCAAGCCCTATGTGTCAACA 1887
|
|
|
Qy 1861 CCTATGGAAGCTACAGTGGCCGAGCAACAGAGTGAAGTGGGGCTACGAGCCCAACG 1920
|
|
|
Db 1888 CCTATGGAAGCTACAGTGGCCGAGCAACAGAGTGAAGTGGGGCTACGAGCCCAACG 1947
|
|
|
Qy 1921 AGGATGGCAGCAGCTGCGTGGG 1942
|
|
|
Db 1948 AGGATGGCAGCAGCTGCGTGGG 1969
|
|
|
```

```
RESULT 4
AK057190
LOCUS
DEFINITION
  AK057190.1 2899 bp mRNA linear PRI 01-AUG-2002
  Homo sapiens cDNA FLJ32628 fis, clone SYN0100018, highly similar
  to Homo sapiens mRNA for ASPIC (acidic secreted protein in
  cartilage) (ASPIC1 gene).
ACCESSION
  AK057190
VERSION
  AK057190.1 GI:16552786
KEYWORDS
  oligo capping, fis (full insert sequence).
SOURCE
  Homo sapiens
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1
  Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
  Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
  Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiya,H.,
  Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
  Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
  Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
  Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
  Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
  Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
  Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
  Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
  Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  NEDO human cDNA sequencing project
  Unpublished
  2 (bases 1 to 2899)
  Isogai,T., Otsuki,T. and Sugiyama,T.
  Direct Submission
  Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
  Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
  (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
  NEDO human cDNA sequencing project supported by Ministry of
  Economy, Trade and Industry of Japan; cDNA full insert sequencing:
  Research Association for Biotechnology (RAB); cDNA library
  construction: Helix Research Institute (HRI) (supported by Japan
  Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
  HRI, and Biotechnology Center, National Institute of Technology and
  Evaluation; clone selection for full insert sequencing: RAB and
  HRI.
FEATURES
  1. .2889
  Location/Qualifiers
  source
```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SYNOV1000118"
/tissue_type="synovial membrane tissue from rheumatoid
arthritis"
/clone_lib="SYNOV1"
/notes="cloning vector: pME18SFL3"

ORIGIN
Query Match      88.8%; Score 1934; DB 9; Length 2889;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1937; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGGAGGCTCAGAGCCAGCCCGGACCGGGCTGGAGCAAGACGAGCGCGCGCGCGCGG 60
|
|
|
Db 235 CGGGAGGCTCAGAGCCAGCCCGGACCGGGCTGGAGCAAGACGAGCGCGCGCGCGG 294
|
|
|
Qy 61 GCAGAGGCGGAGCGAGCGCCCGCTTCCACGCCCCCTAGGGCGGGCGCGAGAGCGGGA 120
|
|
|
Db 295 GCAGAGGCGGAGCGAGCGCCCGCTTCCACGCCCCCTAGGGCGGGCGCGAGAGCGGGA 354
|
|
|
Qy 121 GGAATGGCTCCAGAGCGCTGACCCCGGCACTGTCAGGATGTTACCGTTCCTGCTGCTCT 180
|
|
|
Db 355 GGAATGGCTCCAGAGCGCTGACCCCGGCACTGTCAGGATGTTACCGTTCCTGCTGCTCT 414
|
|
|
Qy 181 GGTTCCTGCCCATCACTCAGAGGGTCCCGAGCGGGCTGAACCCCATGTTCACTGCAGTCACCA 240
|
|
|
Db 415 GGTTCCTGCCCATCACTCAGAGGGTCCCGAGCGGGCTGAACCCCATGTTCACTGCAGTCACCA 474
|
|
|
Qy 241 ACTCAGTTCTGCTCCTGACTATGACAGTAATCCACCAGCGTCAACTATGTTGGGCGAG 300
|
|
|
Db 475 ACTCAGTTCTGCTCCTGACTATGACAGTAATCCACCAGCGTCAACTATGTTGGGCGAG 534
|
|
|
Qy 301 TTACTGATGTGACCAATGATGGGACTTTGAGATGTCGTGCGGGGTACATGAGCCCA 360
|
|
|
Db 535 TTACTGATGTGACCAATGATGGGACTTTGAGATGTCGTGCGGGGTACATGAGCCCA 594
|
|
|
Qy 361 ACCTCGTTCTGAAGTATGACCGGGCCCGAGAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
|
|
|
Db 595 ACCTCGTTCTGAAGTATGACCGGGCCCGAGAGCGGCTGGTGAACATCGCGGTCGATGAGC 654
|
|
|
Qy 421 GCAGCTCACCTTACTACGCGTTCGCGGACCGGAGCGGGAACGCAATGGGGTCAACGCT 480
|
|
|
Db 655 GCAGCTCACCTTACTACGCGTTCGCGGACCGGAGCGGGAACGCAATGGGGTCAACGCT 714
|
|
|
Qy 481 GGCATATGACGGGAGCGCGCGGAGGAGATCTACTTCTCAACCAATATGCTTCT 540
|
|
|
Db 715 GGCATATGACGGGAGCGCGCGGAGGAGATCTACTTCTCAACCAATATGCTTCT 774
|
|
|
Qy 541 CGGGGGTGGCCACGTACACCGCAAGTTGTTCAAAGTTCGGCAATTAACCGTGGGAAGACA 600
|
|
|
Db 775 CGGGGGTGGCCACGTACACCGCAAGTTGTTCAAAGTTCGGCAATTAACCGTGGGAAGACA 834
|
|
|
Qy 601 TCCTCAGCGCATGAGGTCAACGTTGGGCGCGGTGGTGGCCAGCGCTCTTTGGCGGAGCTCTG 660
|
|
|
Db 835 TCCTCAGCGCATGAGGTCAACGTTGGGCGCGGTGGTGGCCAGCGCTCTTTGGCGGAGCTCTG 894
|
|
|
Qy 661 TGGCCCTGTGACAGAAAGGCTCTGGAGCGCTACTCTATCTACATTGCAATTACGCT 720
|
|
|
Db 895 TGGCCCTGTGACAGAAAGGCTCTGGAGCGCTACTCTATCTACATTGCAATTACGCT 954
|
|
|
Qy 721 ACGGTAATCTGGGCGCTGATGCCCTCATTTCAAATGACCTCTGAGGCGAGTGACCTCTCCC 780
|
|
|
Db 955 ACGGTAATCTGGGCGCTGATGCCCTCATTTCAAATGACCTCTGAGGCGAGTGACCTCTCCC 1014
|
|
|
Qy 781 GGGGCAATCTGGCGCTCAGAGATGTGGCTGTGAGCTGGGGTGCAGCAATATACAGGGG 840
|
|
|
Db 1015 GGGGCAATCTGGCGCTCAGAGATGTGGCTGTGAGCTGGGGTGCAGCAATATACAGGGG 1074
|
|
|
Qy 841 GCGGAGCGCTCAGCGTGGGCGCCCATCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
|
|
|
Db 1075 GCGGAGCGCTCAGTGTGGGCGCCCATCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 1134
|
|
|
```



```
QY 901 ATGAGATGGCGCTAACTTCTTTCCACACCGGGCGATGGACCTTTGTGGACGCTG 960
Db 1135 ATGAGATGGCGCTAACTTCTTTCCACACCGGGCGATGGACCTTTGTGGACGCTG 1194
QY 961 CGGCCAGTGTGTGTGGACGACCCACACAGCATGGGCGAGGTGTGSCCTTGGCTGACT 1020
Db 1195 CGGCCAGTGTGTGTGGACGACCCACACAGCATGGGCGAGGTGTGSCCTTGGCTGACT 1254
QY 1021 TCACCGTGTATGGCAAAATGGACATCGTCTATGGCAAACTGGAATGGCCCCACCGCTCT 1080
Db 1255 TCACCGTGTATGGCAAAATGGACATCGTCTATGGCAAACTGGAATGGCCCCACCGCTCT 1314
QY 1081 ATCTGCAAAATGAGCACCCATGGGAGGTCCGCTTCGGGACATCGCTCACCAAGTTCT 1140
Db 1315 ATCTGCAAAATGAGCACCCATGGGAGGTCCGCTTCGGGACATCGCTCACCAAGTTCT 1374
QY 1141 CCATGCCCTCCCTGTCCGACAGCTCATCACGCCGACTTTGACAAATGACACGAGGTGG 1200
Db 1375 CCATGCCCTCCCTGTCCGACAGCTCATCACGCCGACTTTGACAAATGACACGAGGTGG 1434
QY 1201 AGATCTTCTTCAACAAATGCTACCGAGCTCTCTCAGGCAACCGCTCTTCCGCTCA 1260
Db 1435 AGATCTTCTTCAACAAATGCTACCGAGCTCTCTCAGGCAACCGCTCTTCCGCTCA 1494
QY 1261 TCCGTAGAGAGCAGGAGACCCCTCTCATCAGGAGCTCAATCCCGCGACGCTTGGAGC 1320
Db 1495 TCCGTAGAGAGCAGGAGACCCCTCTCATCAGGAGCTCAATCCCGCGAGCTTGGAGC 1554
QY 1321 CTGAGGCGCGGGCACAGGGGGTGTGGTACCGACTTCGACGGAGACGGATGCTGAC 1380
Db 1555 CTGAGGCGCGGGCACAGGGGGTGTGGTACCGACTTCGACGGAGACGGATGCTGAC 1614
QY 1381 TCATCTTGTCCATGAGAGTCCATGGCTCAGCGCTGTCCGTTCCGGGGCAATCAGG 1440
Db 1615 TCATCTTGTCCATGAGAGTCCATGGCTCAGCGCTGTCCGTTCCGGGGCAATCAGG 1674
QY 1441 GCTTCAACAACTGCTGCGAGTGTGCGACGACCCCGGTTTGGGGCTTTGCCAGGG 1500
Db 1675 GCTTCAACAACTGCTGCGAGTGTGCGACGACCCCGGTTTGGGGCTTTGCCAGGG 1734
QY 1501 GAGCTAAGTCTGTCTCTACACCAAGAGTGGGCGCCACCTGAGGATCATCGAGGGG 1560
Db 1735 GAGCTAAGTCTGTCTCTACACCAAGAGTGGGCGCCACCTGAGGATCATCGAGGGG 1794
QY 1561 GCTCAGCTACCTGTGTGAGTGGAGCCGTGGCAACATTTGGCTTGGGAAGGATGAAG 1620
Db 1795 GCTCAGCTACCTGTGTGAGTGGAGCCGTGGCAACATTTGGCTTGGGAAGGATGAAG 1854
QY 1621 CCAGCAGTGTGGAGTGTGCTGCCAGATGGCAAGATGGTGGAGCCGAGCTGGCCAGCG 1680
Db 1855 CCAGCAGTGTGGAGTGTGCTGCCAGATGGCAAGATGGTGGAGCCGAGCTGGCCAGCG 1914
QY 1681 GGGAGTGAACCTCAGTGTGGAGATCCTCTACCCCGGATGAGGACACACTTCAGGACC 1740
Db 1915 GGGAGTGAACCTCAGTGTGGAGATCCTCTACCCCGGATGAGGACACACTTCAGGACC 1974
QY 1741 CAGCCCACTGAGTGTGCCAAGGATTTCCAGCAGGAAATGGCCATTCATGGACA 1800
Db 1975 CAGCCCACTGAGTGTGCCAAGGATTTCCAGCAGGAAATGGCCATTCATGGACA 2034
QY 1801 CCAATGAATGATCCAGTTCCCATTCGTGCTGCCGACAGCCGATGTGTCAACA 1860
Db 2035 CCAATGAATGATCCAGTTCCCATTCGTGCTGCCGACAGCCGATGTGTCAACA 2094
QY 1861 CCTATGGAAGCTACAGGTGCCGGAACCAACAAGATGACGTGGGCTACGAGCCCAAG 1920
Db 2095 CCTATGGAAGCTACAGGTGCCGGAACCAACAAGATGACGTGGGCTACGAGCCCAAG 1942
QY 1921 AGGATGGACACGCTGCTGGG 1942
Db 2155 AGGATGGACACGCTGCTGGG 2176
```

```
RESULT 5
AX427468
LOCUS AX427468 2413 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 41 from Patent WO0123558.
ACCESSION AX427468
VERSION AX427468.1 GI:21537614
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hodgson, D.M., Lincoln, S.E., Russo, F.D., Spiro, P.A., Banville, S.C.,
Bratcher, Shawn, R., Dufour, G.F., Cohen, H.J., Rosen, B.H., Shah, P.,
Chalup, M.S., Hillman, J.L., Jones, A.L., Yu, J.Y., Greenawalt, L.B.,
Panzer, S.R., Roseberry, A.M., Wright, R.J., Chen, W., Liu, T.,
Yap, P.E., Stockdreher, T.K., Amshey, S. and Fong, W.T.
TITLE Human secretory molecules
JOURNAL Patent: WO 0123558-A 41 05-APR-2001;
Incyte Genomics, Inc. (US)
FEATURES
source 1..2413
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 481454.4.dec"
```

## ORIGIN

```
Query Match 87.9%; Score 1913.4; DB 6; Length 2413;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1917; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 19 CCGGGACCGGGGCTGGGAGCAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78
Db 318 CCGGGACCGGGGCTGGGAGCAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 377
QY 79 GCCCGCTTCCACGCGCCCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 138
Db 378 GCCCGCTTCCACGCGCCCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 437
QY 139 ACCCGCGCTTCCAGAGATGTTACCGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 198
Db 438 ACCCGCGCATGTCAGAGATGTTACCGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 497
QY 199 AGGGGTCCCGAGCGGGCTGAACCCATGTTCACTGCAGTCCACCACTCAGTCTGCTCCTG 258
Db 498 AGGGGTCCCGAGCGGGCTGAACCCATGTTCACTGCAGTCCACCACTCAGTCTGCTCCTG 557
QY 259 ACTATGACAGTAAATCCCAACCCAGCTCAACTATGGTGTGGCAGTTACTGATGGAGCATG 318
Db 558 ACTATGACAGTAAATCCCAACCCAGCTCAACTATGGTGTGGCAGTTACTGATGGAGCATG 617
QY 319 ATGGGACCTTTCAGATCGTGTGGCGGGGTACAATGGACCCCAACCTGTTCTGAAGTATG 378
Db 618 ATGGGACCTTTCAGATCGTGTGGCGGGGTACAATGGACCCCAACCTGTTCTGNACTATG 677
QY 379 ACCGGGCGCGAGCGGCTGGTGAACATCGCGGTGATGAGCGGAGCTGACCCCTACTAGC 438
Db 678 ACCGGGCGCGAGCGGCTGGTGAACATCGCGGTGATGAGCGGAGTAAACCCCTACTAGC 737
QY 439 CGCTCGGGACCGGAGGGGAAACGCGCATTTGGGGTCAACAGCTTCGACATCGAGCGGAGCG 498
Db 738 CGCTCGGGACCGGAGGGGAAACGCGCATTCGGGGTCAACAGCTTCGACATCGAGCGGAGCG 797
QY 499 GCGGGAGAGATCTACTTCTTCAACAACAATTAATGCTTCTCGGGGTGGCCAGCTACA 558
Db 798 GCGGGAGAGATCTACTTCTTCAACAACAATTAATGCTTCTCGGGGTGGCCAGCTACA 857
QY 559 CGGACAAGTCTTCAAGTTCGGCAATTAACCGGTGGGAAGACATCTGAGCGATGAGGTCA 618
Db 858 CGGACAAGTCTTCAAGTTCGGCAATTAACCGGTGGGAAGACATCTGAGCGATGAGGTCA 917
```



Db	513		ACCGGCCCAGAAAGCGGCTGGTGAACATCGCGGTGGATGAGCGCAGTAACCCCTACTACG	572
QY	439		CGCTCGGGACCGGACGGGAAACGCCATTGGGGTACAGCTTGGACATCGACGCGGAGC	498
Db	573		CGCTCGGGACCGGACGGGAAACGCCATTGGGGTACAGCTTGGACATCGACGCGGAGC	632
QY	499		GCCGGAGGAGATCTACTTCTCTCAACACCAATAATGCCCTTCTCGGGGTGGCCACGTACA	558
Db	633		GCCGGAGGAGATCTACTTCTCTCAACACCAATAATGCCCTTCTCGGGGTGGCCACGTACA	692
QY	559		CCGACAAGTTGTTCAAGTTCGCCAATAACCGGTGGGAAGACATCTCTGACCGATAGGTC	618
Db	693		CCGACAAGTTGTTCAAGTTCGCCAATAACCGGTGGGAAGACATCTCTGACCGATAGGTC	752
QY	619		ACGTGCCCGCTGGTGTGGCCAGCCCTTTTGGCCGACGCTCTGTGGCCCTGTGTGGACAA	678
Db	753		ACGTGCCCGCTGGTGTGGCCAGCCCTTTTGGCCGACGCTCTGTGGCCCTGTGTGGACAA	812
QY	679		AGGGCTTGGACGCTACTCTATCTACTCATTTGCCAATTACGGCTTACGGTAATGTGGCCCTG	738
Db	813		AGGGCTTGGACGCTACTCTATCTACTCATTTGCCAATTACGGCTTACGGTAATGTGGCCCTG	872
QY	739		ATGCCCTCATTTGAAATGACACCTGAGGCGAGTGACCTCTCCGCGGGCAATCTCGCGCTCA	798
Db	873		ATGCCCTCATTTGAAATGACACCTGAGGCGAGTGACCTCTCCGCGGGCAATCTCGCGCTCA	932
QY	799		GAGATGTGGCTGTGAGGCTGGGGTCAGCAATATACAGGGGGCCGAGGCGCTCAGCGTGG	858
Db	933		GAGATGTGGCTGTGAGGCTGGGGTCAGCAATATACAGGGGGCCGAGGCGCTCAGCGTGG	992
QY	859		GCCCATCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCTTA	918
Db	993		GCCCATCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCTTA	1052
QY	919		TCCTTTCCACAAACGGGGCGATGCGACCTTTGTGGACGCTGCGGCGAGTCTGCTGTGG	978
Db	1053		TCCTTTCCACAAACGGGGCGATGCGACCTTTGTGGACGCTGCGGCGAGTCTGCTGTGG	1112
QY	979		ACGACCCCCACAGCATGGGCGAGGTGTGCGCTCGGCTGACTTCAACCGTATGCGCAAG	1038
Db	1113		ACGACCCCCACAGCATGGGCGAGGTGTGCGCTCGGCTGACTTCAACCGTATGCGCAAG	1172
QY	1039		TGGACATCGTCTATGGCAACTGGAAATGGCCCCCACCGCTCTATCTGCAATGAGCACCC	1098
Db	1173		TGGACATCGTCTATGGCAACTGGAAATGGCCCCCACCGCTCTATCTGCAATGAGCACCC	1232
QY	1099		ATGGGAAGTCCGCTTCCGGGACATCGCTCACCCAGTTCTCCATGCCCCCTCCCTGTCC	1158
Db	1233		ATGGGAAGTCCGCTTCCGGGACATCGCTCACCCAGTTCTCCATGCCCCCTCCCTGTCC	1292
QY	1159		GCAAGGTCTATCACCGCGACTTTGACAATGACAGGAGCTGGAGATCTTCTTCAAGCAAC	1218
Db	1293		GCAAGGTCTATCACCGCGACTTTGACAATGACAGGAGCTGGAGATCTTCTTCAAGCAAC	1352
QY	1219		TTGCGCTACCGAGCTCTCTAGCAACACCGCTCTTTCGCGTCACTCCGTAGAGACACGGAG	1278
Db	1353		TTGCGCTACCGAGCTCTCTAGCAACACCGCTCTTTCGCGTCACTCCGTAGAGACACGGAG	1412
QY	1279		ACCCCTCTACGAGGAGCTCAATCCCGGCGACCGCTTGGAGCTTGGGCGGGGACAG	1338
Db	1413		ACCCCTCTACGAGGAGCTCAATCCCGGCGACCGCTTGGAGCTTGGGCGGGGACAG	1472
QY	1339		GGGGTGTGGTACCGCACTTCCAGGGAGACGGGATGTGTGACCTCATCTTGTCCCATGGAG	1398
Db	1473		GGGGTGTGGTACCGCACTTCCAGGGAGACGGGATGTGTGACCTCATCTTGTCCCATGGAG	1532
QY	1399		AGTCCATGGCTCAGCGCTGTCTCGTCTTCGCGGGCAATCAGGCGTTCAACAACTGGC	1458
Db	1533		AGTCCATGGCTCAGCGCTGTCTCGTCTTCGCGGGCAATCAGGCGTTCAACAACTGGC	1592
QY	1459		TGCGAGTGGTGGCC-ACGACCCCGGTTTGGGGCCCTTTGCCAGGGAGCTAAGTCTGTCTC	1517

Db	1593	TGCGAGTGGTCCAAACGACACCGCGTTTGGGGCCCTTTGCCAGGGGAGACTAAAGTCGTGCTC	1652
QY	1518	TACACCAAGAAGATGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTACTCTGTGT	1577
Db	1653	TACACCAAGAAGATGGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACTCTGTGT	1712
QY	1578	GAGATGGAGCCCGTGGCCACACTTTGGCCCTGGGGAAGNATGAAGCCAGCAGTGTGGAGTG	1637
Db	1713	GAGATGGAGCCCGTGGCCACACTTTGGCCCTGGGGAAGNATGAAGCCAGCAGTGTGGAGTG	1772
QY	1638	ACGTGCCCAGATGGCAAGATGGTCAGCCCGGAACGTGGCCACGCGGGGAGATGAATCAAGT	1697
Db	1773	ACGTGCCCAGATGGCAAGATGGTCAGCCCGGAACGTGGCCACGCGGGGAGATGAATCAAGT	1832
QY	1698	CTGGAGATCTCTTACCCCGCGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT	1757
Db	1833	CTGGAGATCTCTTACCCCGCGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT	1892
QY	1758	GGCCAAAGATTCTCCACAGCAAGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA	1816
Db	1893	GGCCAAAGATTCTCCACAGCAAGAAATGGCCATTG-CATGGACACCAATGAATGCATCCA	1952
QY	1817	GTTCCTCAATCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAAACCTATGGAAGCTACAG	1876
Db	1953	GTTCCTCAATCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAAACCTATGGAAGCTACAG	2012
QY	1877	GTGCCGGAACCAAGAAGTGCAGTCGGGCTCAGAGCCCAACGAGGATGGCACGCCTG	1936
Db	2013	GTGCCGGAACCAAGAAGTGCAGTCGGGCTCAGAGCCCAACGAGGATGGCACGCCTG	2072
QY	1937	CGTGG 1941	
Db	2073	CGTGG 2077	
RESULT 7			
AX357100			
LOCUS	AX357100	2589 bp	DNA linear PAT 13-FEB-2002
DEFINITION	Sequence 1 from Patent WO0206478.		
ACCESSION	AX357100		
VERSION	AX357100.1	GI:18674281	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.		
JOURNAL	Extracellular matrix protein		
	Patent: WO 0206478-A 1 24-JAN-2002;		
FEATURES	Akzo Nobel N.V. (NL)		
source	Location/Qualifiers		
	1..2589		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	83.1%;	Score 1808.8;	DB 6; Length 2589;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 1810;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
QY	131	GAGCGGTGACCCGGCATGTCCAGAGTGTACCGTCTCTGCTGCTGCTCTGTTCTGCC	190
Db	43	GCGGGTGACCCGGCATGTCCAGAGTGTACCGTCTCTGCTGCTCTGTTCTGCC	102
QY	191	CATCACTGAGGGTCCCAAGCGGGCTGAACCCATGTTCACTGCAGTCACCACTCAGTTCT	250
Db	103	CATCACTGAGGGTCCCAAGCGGGCTGAACCCATGTTCACTGCAGTCACCACTCAGTTCT	162
QY	251	GCCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGGTGGCAGTTACTGATGT	310
Db	163	GCCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGGTGGCAGTTACTGATGT	222

Qy	1391	CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTTCTCCGGGGCAATCAGGCTTCAACAA	1450
Db	1303	CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTTCTCCGGGGCAATCAGGCTTCAACAA	1362
Qy	1451	CAACTGGGTGGAGTGGTCCACCGCCGTTTGGGGCTTTCGCGGGAGCTTAAGGT	1510
Db	1363	CAACTGGGTGGAGTGGTCCACCGCCGTTTGGGGCTTTCGCGGGAGCTTAAGGT	1422
Qy	1511	CGTCTCTACACCAAGAGAGTGGGGCCACTTCAGGATCATTCGACGGGGCTCAGGCTA	1570
Db	1423	CGTCTCTACACCAAGAGAGTGGGGCCACTTCAGGATCATTCGACGGGGCTCAGGCTA	1482
Qy	1571	CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCTGGGGAGAGATGAAGCCAGAGTGT	1630
Db	1483	CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCTGGGGAGAGATGAAGCCAGAGTGT	1542
Qy	1631	GGAGTGTGAGTGGCCAGATGGTCAAGATGGTGGAGCCGAGTGGCCAGCGGAGATGAA	1690
Db	1543	GGAGTGTGAGTGGCCAGATGGTCAAGATGGTGGAGCCGAGTGGCCAGCGGAGATGAA	1602
Qy	1691	CTCAGTGTGAGATGCTCTTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCACT	1750
Db	1603	CTCAGTGTGAGATGCTCTTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCACT	1662
Qy	1751	GGAGTGTGGCCAAAGATTTCTCCAGCAGAAATGGCCATTGTCATGGACACCAATGAATG	1810
Db	1663	GGAGTGTGGCCAAAGATTTCTCCAGCAGAAATGGCCATTGTCATGGACACCAATGAATG	1722
Qy	1811	CATCAGTGTCCATTCGTGTGCTTCGAGACAGCCGCTATGTCACACCTATGGAAG	1870
Db	1723	CATCAGTGTCCATTCGTGTGCTTCGAGACAGCCGCTATGTCACACCTATGGAAG	1782
Qy	1871	CTACAGTGTCCGACCAACAGAGTGCAGTCCGGGCTACGAGCCCAACGAGGATGGCAC	1930
Db	1783	CTACAGTGTCCGACCAACAGAGTGCAGTCCGGGCTACGAGCCCAACGAGGATGGCAC	1842
Qy	1931	AGCCTGCGTGGG 1942	
Db	1843	AGCCTGCGTGGG 1854	
RESULT 8			
AX357115			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 1809; Conservative			
Qy			
Db			
Qy			

Qy	311	GGACCATGATGGGACATTTGAGATGCTGCTGGCGGGGTACATGAGACCCACCTGGTTCT	370
Db	223	GGACCATGATGGGACATTTGAGATGCTGCTGGCGGGGTACATGAGACCCACCTGGTTCT	282
Qy	371	GAAGTATGACCGGCGCCAGAGCGGCTGGTGAACATCGCGTTCGATGAGCGCAGCTCACC	430
Db	283	GAAGTATGACCGGCGCCAGAGCGGCTGGTGAACATCGCGTTCGATGAGCGCAGCTCACC	342
Qy	431	CTACTACCGCTGCGGGACCGGACGGGGAACGCCATTTGGGGTTCACAGCTTCGACATCGA	490
Db	343	CTACTACCGCTGCGGGACCGGACGGGGAACGCCATTTGGGGTTCACAGCTTCGACATCGA	402
Qy	491	CGGGAGCGGCGGGAGAGATCTACTTCTCTCAACCAATTAATGCTTCTCGGGGGTGGC	550
Db	403	CGGGAGCGGCGGGAGAGATCTACTTCTCTCAACCAATTAATGCTTCTCGGGGGTGGC	462
Qy	551	CACGTACACGACAGTTGTTCAAGTTCGCGAATAACCGGTGGGAAGACATCCTCAGCGA	610
Db	463	CACGTACACGACAGTTGTTCAAGTTCGCGAATAACCGGTGGGAAGACATCCTCAGCGA	522
Qy	611	TGAGTCAACGTGGCCCGTGGTGTGGCCAGCCCTTTTGGCCGACGCTCTGTGGCCCTGTGT	670
Db	523	TGAGTCAACGTGGCCCGTGGTGTGGCCAGCCCTTTTGGCCGACGCTCTGTGGCCCTGTGT	582
Qy	671	GGACAGAAAGGGCTCTGGACGCTACTTACTACATTTGGCAATTAAGCCCTAGGTAATGT	730
Db	583	GGACAGAAAGGGCTCTGGACGCTACTTACTACATTTGGCAATTAAGCCCTAGGTAATGT	642
Qy	731	GGGCGCTGATGCCCTCATTTGAATGAAATGAGCCCTGAGGCGAGTACCTCTCCGGGGCAATCT	790
Db	643	GGGCGCTGATGCCCTCATTTGAATGAAATGAGCCCTGAGGCGAGTACCTCTCCGGGGCAATCT	702
Qy	791	GGCGCTCAGAGATGGCTGCTGAGGCTGGGGTTCAGCAATATACAGGGGGCCGAGGCT	850
Db	703	GGCGCTCAGAGATGGCTGCTGAGGCTGGGGTTCAGCAATATACAGGGGGCCGAGGCT	762
Qy	851	CAGCGTGGGCGCCATCTCAGCAGAGTGCCTCGGATATCTTCTGGGCAATGAGATGG	910
Db	763	CAGCGTGGGCGCCATCTCAGCAGAGTGCCTCGGATATCTTCTGGGCAATGAGATGG	822
Qy	911	GCTAACTTCTTTTCCAAACCGGGCGATGGACCTTTGTGGAGCTGGCGCCAGTGC	970
Db	823	GCTAACTTCTTTTCCAAACCGGGCGATGGACCTTTGTGGAGCTGGCGCCAGTGC	882
Qy	971	TGGTGTGAGACGACCCCGACAGCATGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA	1030
Db	883	TGGTGTGAGACGACCCCGACAGCATGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA	942
Qy	1031	TGGCAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCGCCAGCGCTCTATCTGCAAT	1090
Db	943	TGGCAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCGCCAGCGCTCTATCTGCAAT	1002
Qy	1091	GAGCACCATGGGAAGTCCGCTTCCGGGACATCGCTCACCACCAAGTTCCTCCATGCCCTC	1150
Db	1003	GAGCACCATGGGAAGTCCGCTTCCGGGACATCGCTCACCACCAAGTTCCTCCATGCCCTC	1062
Qy	1151	CCCTGTCCGACGCTCATCACCGCGACTTTGCAATGACCGAGCTGGAGATCTTCTT	1210
Db	1063	CCCTGTCCGACGCTCATCACCGCGACTTTGCAATGACCGAGCTGGAGATCTTCTT	1122
Qy	1211	CACACATTTGCTACCGAGCTCTCAGCAACCGCTTTCGGGTGATCCGATGAGA	1270
Db	1123	CACACATTTGCTACCGAGCTCTCAGCAACCGCTTTCGGGTGATCCGATGAGA	1182
Qy	1271	GCACGAGACCCCTCATCGAGGAGTCAATCCCGCGAGCGCTTGGAGCTCAGGCGCG	1330
Db	1183	GCACGAGACCCCTCATCGAGGAGTCAATCCCGCGAGCGCTTGGAGCTCAGGCGCG	1242
Qy	1331	GGGACAGGGGGTGTGACCGGCTTCGACGAGAGCGGATGCTGACCTCATCTGTCT	1390
Db	1243	GGGACAGGGGGTGTGACCGGCTTCGACGAGAGCGGATGCTGACCTCATCTGTCT	1302

Db 103 CATCACTGAGGGTCCACAGGGCTGAACCCATGTTCTAGAGTCAACAACTAGTTCT 162  
QY 251 GCCTCCTGACTATGACAGTAATCCACCGAGTCAACTATGTTGGAGTTACTGATGT 310  
Db 163 GCCTCCTGACTATGACAGTAATCCACCGAGTCAACTATGTTGGAGTTACTGATGT 222  
QY 311 GGACCATATGGGACTTTTGAGATCGTGGGGGTGACATGAGACCAACCTGGTTCT 370  
Db 223 GGACCATATGGGACTTTTGAGATCGTGGGGGTGACATGAGACCAACCTGGTTCT 282  
QY 371 GAAATATGACCGGCCCCAGAAAGCGCTGGTGAACATCGCGTTCATGAGCGCAGCTCACC 430  
Db 283 GAAATATGACCGGCCCCAGAAAGCGCTGGTGAACATCGCGTTCATGAGCGCAGCTCACC 342  
QY 431 CTACTAGCGCTGCGGACCGGACCGGAGGGAACCGCATTTGGGTTCACAGCTCGCATCGA 490  
Db 343 CTACTAGCGCTGCGGACCGGACCGGAGGGAACCGCATCGGGGTTCACAGCTCGCATCGA 402  
QY 491 CGGGACCGCGCGGAGGAGATCTACTTCCCTCAACACCAATAATGCCCTTCTCGGGGGTGGC 550  
Db 403 CGGGACCGCGCGGAGGAGATCTACTTCCCTCAACACCAATAATGCCCTTCTCGGGGGTGGC 462  
QY 551 CAGCTACACCGCAAGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCTCGAGCGA 610  
Db 463 CAGCTACACCGCAAGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCTCGAGCGA 522  
QY 611 TGAGGTCAACGTTGGCCCGTGGTGGCCAGCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670  
Db 523 TGAGGTCAACGTTGGCCCGTGGTGGCCAGCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582  
QY 671 GGACAGAAAGGGCTCTGGACGCTACTTATCTACATTGGCAATTACGCCCTACGGTAATGT 730  
Db 583 GGACAGAAAGGGCTCTGGACGCTACTTATCTACATTGGCAATTACGCCCTACGGTAATGT 642  
QY 731 GGGCCCTGATGCCCCTCATTTGAATGGAACCTGAGGCCAGTGACCTCTCCCGGGGCAATCT 790  
Db 643 GGGCCCTGATGCCCCTCATTTGAATGGAACCTGAGGCCAGTGACCTCTCCCGGGGCAATCT 702  
QY 791 GGGGCTCAGAGATGGCTGCTGAGGCTGGGGTCAAGCAATATACAGGGGGCCGAGGCGT 850  
Db 703 GGGGCTCAGAGATGGCTGCTGAGGCTGGGGTCAAGCAATATACAGGGGGCCGAGGCGT 762  
QY 851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTTCGACAAATGAGATGG 910  
Db 763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTTCGACAAATGAGATGG 822  
QY 911 GCGTAACCTTCCTTTCCCAACCGGGGCGATGGACCTTTGTGGACGCTGGGCCAGTGC 970  
Db 823 GCGTAACCTTCCTTTCCCAACCGGGGCGATGGACCTTTGTGGACGCTGGGCCAGTGC 882  
QY 971 TGGTGTGGACGACCCCCACACGATGGGAGTGTGCGGCTGGCTGATTCACACCGTGA 1030  
Db 883 TGGTGTGGACGACCCCCACACGATGGGAGTGTGCGGCTGGCTGATTCACACCGTGA 942  
QY 1031 TGGCAAGTGGACATCGTCTATGGCACTGGAATGGCCGCCACCGCTCTATCTGCAAT 1090  
Db 943 TGGCAAGTGGACATCGTCTATGGCACTGGAATGGCCGCCACCGCTCTATCTGCAAT 1002  
QY 1091 GAGCACCATGGGAAGTCCGCTTCGGGACATCGCTCACCACCAAGTTCTCCATGCCCTC 1150  
Db 1003 GAGCACCATGGGAAGTCCGCTTCGGGACATCGCTCACCACCAAGTTCTCCATGCCCTC 1062  
QY 1151 CCCTGTCCGACGGTTCATCAGCCCGAATTTGACAAATGACAGGAGCTGGAGATCTCTT 1210  
Db 1063 CCCTGTCCGACGGTTCATCAGCCCGAATTTGACAAATGACAGGAGCTGGAGATCTCTT 1122  
QY 1211 CAACAACATTGCTACCGCAGCTCCTCAGCCAAACCGCTCTTCGGGTATCCGATAGAGA 1270  
Db 1123 CAACAACATTGCTACCGCAGCTCCTCAGCCAAACCGCTCTTCGGGTATCCGATAGAGA 1182  
QY 1271 GCACGGAGACCCCTCATCTGAGGAGCTCAATCCCGGCGAGCTTGGAGCCTGAGGGCG 1330

Db 1183 GCACGGAGACCCCTCATCTGAGGAGCTCAATCCCGGCGACGCTTGGAGCCTGAGGGCG 1242  
QY 1331 GGGCACAGGGGCTGTGGTGACCGCTTCGACGGAGACGGGATGCTGGACCTCATCTTGTTC 1390  
Db 1243 GGGCACAGGGGCTGTGGTGACCGCTTCGACGGAGACGGGATGCTGGACCTCATCTTGTTC 1302  
QY 1391 CCATGGAGAGTCCATGCTCAGCGCTGTCCGCTTCCCGGGGCAATCAGGGGCTTCAACAA 1450  
Db 1303 CCATGGAGAGTCCATGCTCAGCGCTGTCCGCTTCCCGGGGCAATCAGGGGCTTCAACAA 1362  
QY 1451 CAACGTGCTGCGAGTGTGTCACGACCCCGTGTGGGCTTTGGCCAGGGGAGCTTAAGGT 1510  
Db 1363 CAACGTGCTGCGAGTGTGTCACGACCCCGTGTGGGCTTTGGGCTTTGCGAGGGAGCTTAAGGT 1422  
QY 1511 CGTCTCTTACACCAAGAGAGTGGGGCCCACTCTCAGGATCATCGACGGGGCTCAGGCTA 1570  
Db 1423 CGTCTCTTACACCAAGAGAGTGGGGCCCACTCTCAGGATCATCGACGGGGCTCAGGCTA 1482  
QY 1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCTTGGGGAAGATGAAGCCAGCAGTGT 1630  
Db 1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCTTGGGGAAGATGAAGCCAGCAGTGT 1542  
QY 1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCGCGAAACGTGGCCAGCGGGAGATGAA 1690  
Db 1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCGCGAAACGTGGCCAGCGGGAGATGAA 1602  
QY 1691 CTGAGTCTGGAGATCCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750  
Db 1603 CTGAGTCTGGAGATCCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662  
QY 1751 GGAGTGTGGCCAAAGGATTTCCAGCAGGAGAAATGGCCATTGATGACACCAATGAATG 1810  
Db 1663 GGAGTGTGGCCAAAGGATTTCCAGCAGGAGAAATGGCCATTGATGACACCAATGAATG 1722  
QY 1811 CATCCAGTTCCTATGTTGTCCTTCAGACAAAGCCGATATGTTCAACACCTATGGAAG 1870  
Db 1723 CATCCAGTTCCTATGTTGTCCTTCAGACAAAGCCGATATGTTCAACACCTATGGAAG 1782  
QY 1871 CTACAGTGTCCGACCAACAAAGAGTGCAGTCCGGGCTAGAGCCCAACAGGATGGCAC 1930  
Db 1783 CTACAGTGTCCGACCAACAAAGAGTGCAGTCCGGGCTAGAGCCCAACAGGATGGCAC 1842  
QY 1931 AGCCTCGCTGG 1941  
Db 1843 AGCCTCGCTGG 1853

RESULT 9  
AX357117  
LOCUS  
DEFINITION  
Sequence 18 from Patent WO0206478.  
ACCESSION  
AX357117  
VERSION  
AX357117.1 GI:18674297  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
Heus, H.C., Nelissen, R.L. and Meeuwisse, C.M.  
Extracellular matrix protein  
Patent: WO 0206478-A 18 24-JAN-2002;  
Akzo Nobel N.V. (NL)  
FEATURES  
Location/Qualifiers  
1..2145  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 83.0%; Score 1807.8; DB 6; Length 2145;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1809; Conservative 0; Mismatches 2; Indels 0; Gaps 0;





ORIGIN	/db_xref="taxon:9606"									
Query Match	82.4%;	Score 1794.4;	DB 6;	Length 1959;						
Best Local Similarity	99.9%;	Pred.No. 0;								
Matches 1795;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;						
QY	147	ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGCTTTCTGCCATCACTAGAGGGTTC	206							
Db	1	ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGCTTTCTGCCATCACTAGAGGGTTC	60							
QY	207	CAGCGGGCTGAACCCATGTTCTACTGAGTCAACCACTCAAGTTCGTCTCTGACTATGAC	266							
Db	61	CAGCGGGCTGAACCCATGTTCTACTGAGTCAACCACTCAAGTTCGTCTCTGACTATGAC	120							
QY	267	AGTAATCCACCACCGAGCTCAACTATGCTGGGAGTTACTGATGTGACCATGATGGGAC	326							
Db	121	AGTAATCCACCACCGAGCTCAACTATGCTGGGAGTTACTGATGTGACCATGATGGGAC	180							
QY	327	TTTGAGATCGTCTGGGGGTCAATGACCAACCTGGTTCTGAAGTATCACCGGGC	386							
Db	181	TTTGAGATCGTCTGGGGGTCAATGACCAACCTGGTTCTGAAGTATCACCGGGC	240							
QY	387	CAGAAAGCGGCTGCTGAACATCGCGGTGCATGAGCGAGCTCAACCTACTACCGGTGGG	446							
Db	241	CAGAAAGCGGCTGCTGAACATCGCGGTGCATGAGCGAGCTCAACCTACTACCGGTGGG	300							
QY	447	GACCGGCAGGGGAACGCCATGCGGTTCACAGCCTCGACATCGACGGGAGCGCCGGGAG	506							
Db	301	GACCGGCAGGGGAACGCCATGCGGTTCACAGCCTCGACATCGACGGGAGCGCCGGGAG	360							
QY	507	GAGATCTACTTTCCTCAACACCAATTAATGCTTCTCGGGGTGGCCAGCTACACCGACAAG	566							
Db	361	GAGATCTACTTTCCTCAACACCAATTAATGCTTCTCGGGGTGGCCAGCTACACCGACAAG	420							
QY	567	TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTTGAGGGATGAGGTCAACGTGCC	626							
Db	421	TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTTGAGGGATGAGGTCAACGTGCC	480							
QY	627	CGTGGTGTGGCAGCCCTTTTGGCCGAGCGCTGTGGCCTCTGTGGACAGAAAGGGCTCT	686							
Db	481	CGTGGTGTGGCAGCCCTTTTGGCCGAGCGCTGTGGCCTCTGTGGACAGAAAGGGCTCT	540							
QY	687	GGACGCTACTCTATCTACATGCGCAATPACGCCTAAGGTAAATGTGGGCCCTGATGCCCTC	746							
Db	541	GGACGCTACTCTATCTACATGCGCAATPACGCCTAAGGTAAATGTGGGCCCTGATGCCCTC	600							
QY	747	ATTGAAATGACCTGAGGCGAGTGACCTCTCCGGGGCATTCCTGGCGCTCAGAGATGTG	806							
Db	601	ATTGAAATGACCTGAGGCGAGTGACCTCTCCGGGGCATTCCTGGCGCTCAGAGATGTG	660							
QY	807	GCTGCTGAGGCTGGGGTCAGCAATAATACAGGGGGCCAGGCGCTCAGCGTGGGCCCCATC	866							
Db	661	GCTGCTGAGGCTGGGGTCAGCAATAATACAGGGGGCCAGGCGCTCAGCGTGGGCCCCATC	720							
QY	867	CTCAGCAGCAGTGCCTCGGATATCTTCTGCGAATAATGAGAAATGGGCTTAACTCTCTTTTC	926							
Db	721	CTCAGCAGCAGTGCCTCGGATATCTTCTGCGAATAATGAGAAATGGGCTTAACTCTCTTTTC	780							
QY	927	CACAAACGGGGCGATGSCACTTTTGTGGACGCTCGGCCAGTCTGTGTGGACGACCCC	986							
Db	781	CACAAACGGGGCGATGSCACTTTTGTGGACGCTCGGCCAGTCTGTGTGGACGACCCC	840							
QY	987	CACAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGCCAAAGTGGACATC	1046							
Db	841	CACAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGCCAAAGTGGACATC	900							
QY	1047	GTCATATGGCAACTGGAATGSCCCCCCAGCGCTCTATCTGCAATGAGCAACCCATGGGAAG	1106							
Db	901	GTCATATGGCAACTGGAATGSCCCCCCAGCGCTCTATCTGCAATGAGCAACCCATGGGAAG	960							
QY	1107	GTCCGGCTTCGGGACATCGCCTCACCCAGTTCTCCATGCCCTCCCTGTCCGACGGTCT	1166							



ORIGIN		/organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"	
Query Match		82.4%; Score 1794.4; DB 6; Length 2507;	
Best Local Similarity		99.9%; Pred. No. 0;	
Matches 1795; Conservative		0; Mismatches 1; Indels 0; Gaps 0;	
QY	1107	GTCCGCTCCGGGACATGCGCTCACCAGTTCTCCATGCCCTCCCTGTCCGACGGTC	1166
Db	961	GTCCGCTCCGGGACATGCGCTCACCAGTTCTCCATGCCCTCCCTGTCCGACGGTC	1020
QY	1167	ATCACCGCGGACTTTGACAAATGACAGAGCTGGAGATCTTCTTCAACAACATTCCTTAC	1226
Db	1021	ATCACCGCGGACTTTGACAAATGACAGAGCTGGAGATCTTCTTCAACAACATTCCTTAC	1080
QY	1227	CGCAGCTCCTCAGGCCAACCGCTCTTCCGCGCTCATCCCTAGAGACA CGGAGACCCCTC	1286
Db	1081	CGCAGCTCCTCAGGCCAACCGCTCTTCCGCGCTCATCCCTAGAGACA CGGAGACCCCTC	1140
QY	1287	ATCGAGGAGCTCAATCCCGGGAACCGCTTGAGAGCTGAGGGCGGGGACAGGGGTG	1346
Db	1141	ATCGAGGAGCTCAATCCCGGGAACCGCTTGAGAGCTGAGGGCGGGGACAGGGGTG	1200
QY	1347	GTGACCGGACTTCGACGAGAGCGGATGCTGGACCTCATCTTCTCCATGGAGAGTCCATG	1406
Db	1201	GTGACCGGACTTCGACGAGAGCGGATGCTGGACCTCATCTTCTCCATGGAGAGTCCATG	1260
QY	1407	GCTCAGCGCTGTCCGCTCTTCCGGGCAATCAGGGCTTCAACAACATTCGGAGTG	1466
Db	1261	GCTCAGCGCTGTCCGCTCTTCCGGGCAATCAGGGCTTCAACAACATTCGGAGTG	1320
QY	1467	GTGCAACGACCCCGTTTGGGCTTTCAGGGGAGCTAAGTCTGCTCTACACCAAG	1526
Db	1321	GTGCAACGACCCCGTTTGGGCTTTCAGGGGAGCTAAGTCTGCTCTACACCAAG	1380
QY	1527	AGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGGCTACCTGTGAGATGGAG	1586
Db	1381	AGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGGCTACCTGTGAGATGGAG	1440
QY	1587	CCGCTGGCACACTTTGGCTTGGGAGAGTGAAGCCAGAGTGTGAGGTGAGCTGGGCA	1646
Db	1441	CCGCTGGCACACTTTGGCTTGGGAGAGTGAAGCCAGAGTGTGAGGTGAGCTGGGCA	1500
QY	1647	GATGGCAAGTGTGAGCGGAAAGTGGCCAGCGGGAGTGAACCTCAGTGTGGAGATC	1706
Db	1501	GATGGCAAGTGTGAGCGGAAAGTGGCCAGCGGGAGTGAACCTCAGTGTGGAGATC	1560
QY	1707	CTTACCCCGGATCAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGCCAAGGA	1766
Db	1561	CTTACCCCGGATCAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGCCAAGGA	1620
QY	1767	TTCTCCAGCAGCAAAATGSCCATTCATGGACACCAATGAATGCATCCAGTTCCTATC	1826
Db	1621	TTCTCCAGCAGCAAAATGSCCATTCATGGACACCAATGAATGCATCCAGTTCCTATC	1680
QY	1827	GTGTGCCCTCGACAGACCCCGTATGTGTCAACACCTATGAAAGTACAGTGTCCGAC	1886
Db	1681	GTGTGCCCTCGACAGACCCCGTATGTGTCAACACCTATGAAAGTACAGTGTCCGAC	1740
QY	1887	AACAAGAGTGCAGTGGGGCTACGAGCCCAAGAGGATGGACAGCTCGTGGG	1942
Db	1741	AACAAGAGTGCAGTGGGGCTACGAGCCCAAGAGGATGGACAGCTCGTGGG	1796
RESULT 12		HSA279016 2507 bp mRNA linear PRI 11-MAR-2001	
HSA279016		Homo sapiens mRNA for chondrocyte expressed protein 68 kDa (CEP-68	
LOCUS		gene).	
DEFINITION		AJ279016 GI:10178882	
ACCESSION		CEP-68 gene; chondrocyte expressed protein 68 kDa.	
VERSION		Homo sapiens (human)	
KEYWORDS		Homo sapiens	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		Steck, E., Benz, K., Loreuz, H., Loew, M., Gress, T. and Richter, W.	
TITLE		Chondrocyte expressed protein-68 (CEP-68), a novel human marker	
		gene for cultured chondrocytes	











This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 11:55:50 ; Search time 821 Seconds  
(without alignments)  
11264.710 Million cell updates/sec

Title: US-09-914-958b-35

Perfect score: 2177

Sequence: 1 cggagagctcgagccagcc.....tctcttgaaaaa 2177

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001as.\*  
5: Geneseq2001bs.\*  
6: Geneseq2002s.\*  
7: Geneseq2003as.\*  
8: Geneseq2003bs.\*  
9: Geneseq2003cs.\*  
10: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177	100.0	2177	3	AA75118
2	2043.8	93.9	2089	6	AA229792
3	1913.4	87.9	2413	4	AA501551
4	1893	87.0	2263	4	AAH23100
5	1808.8	83.1	2589	6	AA229777
6	1807.8	83.0	2039	6	AA229793
7	1807.8	83.0	2145	6	AA229793
8	1794.4	82.4	1959	6	AA47918
9	1794.4	82.4	2507	6	AA47918
10	1794.4	82.4	2507	7	ACC50982
11	1793.4	82.4	1947	6	AA47918
12	1687.8	77.5	1849	6	AA562301
13	1408.8	64.7	2178	4	AAH14000
14	1408.8	64.7	2178	7	ACC50981
15	1408.8	62.0	1574	4	AA160925
16	1349	62.0	1501	4	AA160925
17	1345.8	61.8	1501	8	AD849122
18	1345.8	61.8	1501	8	AD849122
19	1025.6	47.1	1143	3	AA265257
20	1025.6	47.1	1143	7	ADA56306
21	1025.6	47.1	1143	7	ADA40142
22	1025.6	47.1	1143	9	AD811646
23	1025.6	47.1	1143	9	AD811646

24	747.8	34.4	883	6	ABL42156
25	687.2	31.6	4794	7	ACC50983
26	687.2	31.6	4794	7	ABX76366
27	556.8	25.6	608	4	AAH05239
28	429	19.7	465	8	ACH31964
29	397.4	18.3	447	8	ACH42110
30	228.4	10.5	319	2	AAT25375
31	191.2	8.8	1400	5	AA83195
32	101	4.6	1216	6	ABQ28697
33	101	4.6	1216	6	ABQ28696
34	72.8	3.3	1216	6	ABQ28694
35	72.8	3.3	1216	6	ABQ28695
36	61.4	2.8	2000	7	ADA71938
37	60	2.8	60	6	ABN43600
38	58.2	2.7	110000	4	AA199683
39	57	2.6	57	6	AA47919
40	55.6	2.6	2000	7	ADA71938
41	54.8	2.5	110000	4	AA199682
42	49.2	2.3	10732	3	AAA10594
43	48.6	2.2	985	6	ABQ44816
44	48.6	2.2	985	6	ABQ44817
45	48.2	2.2	1815	5	AA09400

## ALIGNMENTS

RESULT 1  
AA75118  
ID AAA75118 standard; cDNA; 2177 BP.  
XX  
AC AAA75118;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE cDNA encoding a human secretory protein.  
XX  
KW Human; secretory protein; HSECP; cancer; gastrointestinal disorder;  
KW inflammation; cardiovascular disorder; neurological disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 123..2003  
FT /\*tag= a  
FT /\*product= "secretory protein"  
FT sig\_peptide 123..203  
FT /\*tag= b  
XX  
PN WO200052151-A2.  
XX  
PD 08-SEP-2000.  
XX  
PF 03-MAR-2000; 2000WO-US005621.  
XX  
PR 05-MAR-1999; 99US-0123117P.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Tang YT, Lal P, Baughn MR, Yue H, Au-Young J, Lu DAM, Azimzai Y;  
XX  
XX WPI; 2000-579282/54.  
XX  
XX P-PSDB; AAB08864.  
XX  
XX Twenty two human secretory proteins for diagnosing, treating and  
XX preventing cancer, inflammation, and gastrointestinal, cardiovascular and  
XX neurological disorders.  
XX  
XX Claim 4; Page 102-103; 107pp; English.  
XX  
XX The present sequence encodes a human secretory protein, designated HSECP-  
XX 1. The specification also describes HSECP-2 to HSECP-22. The proteins are  
XX useful for diagnosing, treating and preventing cancer, inflammation, and





Db 943 TGGCAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCAGCCCTCTATCTGCAAT 1002  
Qy 1091 GAGCACCATGGAGGTCCGCTTCCGGGACATCGCTCACCCAGTTCTCCATGCCCTC 1150  
Db 1003 GAGCACCATGGAGGTCCGCTTCCGGGACATCGCTCACCCAGTTCTCCATGCCCTC 1062  
Qy 1151 CCCTGTCCGACGGTCTATCCCGCGACTTTGACAAATGACGAGCTGGAGATCTTCT 1210  
Db 1063 CCCTGTCCGACGGTCTATCCCGCGACTTTGACAAATGACGAGCTGGAGATCTTCT 1122  
Qy 1211 CAACAACATTCCTACCGAGTCTCAGCAACCGCTCTTCCGCTCATCCGTAAGA 1270  
Db 1123 CAACAACATTCCTACCGAGTCTCAGCAACCGCTCTTCCGCTCATCCGTAAGA 1182  
Qy 1271 GCAGGAGACCCCTCATCGAGAGCTCAATCCCGGCGACGCTTGGAGCCTGAGGGCG 1330  
Db 1183 GCAGGAGACCCCTCATCGAGAGCTCAATCCCGGCGACGCTTGGAGCCTGAGGGCG 1242  
Qy 1331 GGGCACAGGGGTGTGGTGACCGACTTCGACGAGACGGGATGTGGACCTCATCTTGT 1390  
Db 1243 GGGCACAGGGGTGTGGTGACCGACTTCGACGAGACGGGATGTGGACCTCATCTTGT 1302  
Qy 1391 CAATGGAGTCCATGGCTCAGCGCTGTCCTCTTCCGGGCAATCAGGGCTTCAAXA 1450  
Db 1303 CAATGGAGTCCATGGCTCAGCGCTGTCCTCTTCCGGGCAATCAGGGCTTCAAXA 1362  
Qy 1451 CAATGGCTCGAGTGGTCCACGACCGCGTGTGGGGCTTTGCCAGGGGAGCTAAGT 1510  
Db 1363 CAATGGCTCGAGTGGTCCACGACCGCGTGTGGGGCTTTGCCAGGGGAGCTAAGT 1422  
Qy 1511 CCGTCTCTACCAAGAGAGTGGGGCCCACTGAGGATCATCGAGGGGCTCAGGCTA 1570  
Db 1423 CCGTCTCTACCAAGAGTGGGGCCCACTGAGGATCATCGAGGGGCTCAGGCTA 1482  
Qy 1571 CCGTGTGAGTGGAGCCCGTGGCACACTTGGCTGGGGAGGATGAAGCCAGCAGTGT 1630  
Db 1483 CCGTGTGAGTGGAGCCCGTGGCACACTTGGCTGGGGAGGATGAAGCCAGCAGTGT 1542  
Qy 1631 GGAGGTGAGTGGCCAGATGGCAAGATGGTGGCCGGAAGCTGGCCAGGGGAGATGA 1690  
Db 1543 GGAGGTGAGTGGCCAGATGGCAAGATGGTGGCCGGAAGCTGGCCAGGGGAGATGA 1602  
Qy 1691 CTCAGTGTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACT 1750  
Db 1603 CTCAGTGTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACT 1662  
Qy 1751 GGAGTGGCCAGGATCTCCAGCAGGAAATGGCATTGATGACACCAATGAATG 1810  
Db 1663 GGAGTGGCCAGGATCTCCAGCAGGAAATGGCATTGATGACACCAATGAATG 1722  
Qy 1811 CATCCAGTTCCTGATTCGCTGTCCTCGAGACAGCCCGTATGTGTCAACACCTATGAAG 1870  
Db 1723 CATCCAGTTCCTGATTCGCTGTCCTCGAGACAGCCCGTATGTGTCAACACCTATGAAG 1782  
Qy 1871 CTCAGTGTGGAGACCAAGAGTGGCTGAGGCTGAGCCCAACAGGATGGCAC 1930  
Db 1783 CTCAGTGTGGAGACCAAGAGTGGCTGAGGCTGAGCCCAACAGGATGGCAC 1842  
Qy 1931 AGCTGTGGCTGGTGGAGCCCTGTTGAAGATAGTGACACCAAGTTGGGAAGAG 1990  
Db 1843 AGCTGTGGCTGGTGGAGCCCTGTTGAAGATAGTGACACCAAGTTGGGAAGAG 1902  
Qy 1991 CCTTGTTCCTGATTCATCTGATCACTGCTTGAATCAGCCCTGGGAATACCTTGTATC 2050  
Db 1903 CCTTGTTCCTGATTCATCTGATCACTGCTTGAATCAGCCCTGGGAATACCTTGTATC 1962  
Qy 2051 AGGAACACTTACCTGGAACTTCTGAGCAGATACAACTCTTATTTAGCTATT 2110  
Db 1963 AGGAACACTTACCTGGAACTTCTGAGCAGATACAACTCTTATTTAGCTATT 2022  
Qy 2111 AATACATTAAAGATTGGGGGTGCTACCTTACATAATAAATCCCAATTTCTCTTGA 2170  
Db 2023 AATACATTAAAGATTGGGGGTGCTACCTTACATAATAAATCCCAATTTCTCTTGA 2082

Qy 2171 AAAAAA 2177  
Db 2083 AAAAAA 2089  
RESULT 3  
ID AAS01551 standard; cDNA; 2413 BP.  
AC AAS01551;  
XX 18-JUL-2001 (first entry)  
XX Human secretory molecule cDNA sptm #41.  
XX Human; secretory molecule; sptm; SPTM; library screening; gene therapy;  
KW cell signalling; cell proliferative disorder; atherosclerosis; cancer;  
KW immune system disease; AIDS; neurological disorder; Alzheimer's disease;  
KW nervous system disease; mental retardation; developmental disorder;  
KW neuromuscular disorder; microarray; microarray; Incyte ID number 4814544dec; ss.  
XX Homo sapiens.  
XX WO200123558-A2.  
XX 05-APR-2001.  
XX 19-SEP-2000; 2000WO-US025610.  
XX 28-SEP-1999; 99US-0156624P.  
PR 28-SEP-1999; 99US-0156625P.  
PR 02-DEC-1999; 99US-0168611P.  
PR 02-DEC-1999; 99US-0168613P.  
PR 02-DEC-1999; 99US-0168614P.  
XX (INCYTE) INCYTE GENOMICS INC.  
XX Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;  
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;  
PI Hillman JL, Jones AL, Yu JY, Greenwalt LB, Panzer SR, Roseberry AM;  
PI Wright RJ, Chen W, Liu TF, Yap PE, Stockreher TK, Amshey S;  
PI Fong WT;  
XX WPI; 2001-258134/26.  
XX New secretory polynucleotides (SPTM) and the polypeptides they encode,  
for use in inducing antibodies and screening libraries of compounds.  
XX Claim 1; Page 146; 161pp; English.  
XX The present sequence for human secretory molecule cDNA sptm #41 (Incyte  
ID number 4814544dec) is 1 of 63 novel sptm cDNA sequences (AAS01511-  
AAS01573) which encode for the secretory polypeptides SPTM. The sptm  
polynucleotides are useful for screening a compound for effectiveness in  
altering expression of a target polynucleotide, where the target  
polynucleotide comprises sptm. Sptm is also useful in a method for  
assessing the toxicity of a test compound. Sptm and its fragments or  
complementary sequences, may be used to identify the presence of and/or  
determine the degree of similarity between two nucleic acid sequences.  
Sptm can also be used for a variety of diagnostic and therapeutic  
purposes, e.g. diagnosing a particular condition, disease or disorder  
associated with cell signalling, such conditions include cell  
proliferative disorders such as atherosclerosis, and cancers including  
leukemia, an immune system disorder e.g. acquired immunodeficiency  
syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's  
disease, nutritional and metabolic disorders of the nervous system, mental  
retardation and other developmental disorders, and muscular dystrophy and  
other neuromuscular disorders. Sptm can also be used to design probes of  
useful in diagnostic assays, which may be used to monitor the progress of  
conditions or disorders associated with abnormal levels of expression of  
sptm. In addition sptm encoding SPTM may be used for somatic or germline  
gene therapy, for inducing antibodies, or in microarrays

XX	SQ	Sequence 2413 BP; 511 A; 714 C; 722 G; 459 T; 0 U; 7 Other;
		Query Match      87.9%;    Score 1913.4;    DB 4;    Length 2413;
		Best Local Similarity    99.7%;    Pred. No. 0;
		Matches 1917;    Conservative    0;    Mismatches    6;    Indels    0;    Gaps    0;
QY	19	CCCGGACCGGGCTGGAGCAAGCAGCGCGCGCGCGCGCGGCGGCGAGAGCGGCAGCGAGC 78
Dd	318	CCCGGACCGGGCTGGAGCAAGCAGCGCGCGCGCGCGCGGCGGCGAGAGCGGCAGCGAGC 377
QY	79	GCCCGCTTCCCACGCCCTTAGCGCGCGGGCCGAGCGGGAGGATGGCTCCGAGCGGTG 138
Dd	378	GCCCGCTTCCCACGCCCTTAGCGCGCGGGCCGAGCGGGAGGATGGCTCCGAGCGGTG 437
QY	139	ACCCCGCATGTCAGAGGATGTTACCGTTTCCTGCTGCTGCTCTGTTCTTGCCCCATCACTG 198
Dd	438	ACCCCGCATGTCAGGAAGTTAACGGTTCCTGCTGCTGCTCTGTTCTTGCCCCATCACTG 497
QY	199	AGGGTCCACGCGGCTGAACCCCATGTTCACTGCACTCACCACTCACTGCACTCACTCACTG 258
Dd	498	AGGGTCCACGCGGCTGAACCCCATGTTCACTGCACTCACCACTCACTGCACTCACTCACTG 557
QY	259	ACTATGACAGTAATCCACCACCGACTAACTATATGTTGGCGAGTTACTGATGTGACCATG 318
Dd	558	ACTATGACAGTAATCCACCACCGACTAACTATATGTTGGCGAGTTACTGATGTGACCATG 617
QY	319	ATGGGACTTTGAGATCGTCTGCGGGGTACAATGGACCAACCTGGTTCTGTAAGTATG 378
Dd	618	ATGGGACTTTGAGATCGTCTGCGGGGTACAATGGACCAACCTGGTTCTGTAAGTATG 677
QY	379	ACGGGCCCGAAGCGCTGGTGAACATCGCGTTCGATGAGCGCAGCTCACCTTACTACG 438
Dd	678	ACGGGCCCGAAGCGCTGGTGAACATCGCGTTCGATGAGCGCAGTAAACCCCTTACTACG 737
QY	439	CGCTGCGGACCGGCAGGGAAACCGATTGGGGTCAAGCTGTCGACATCGACGGGACG 498
Dd	738	CGCTGCGGACCGGCAGGGAAACCGATTGGGGTCAAGCTGTCGACATCGACGGGACG 797
QY	499	GCGCGGAGGAGATCTACTTCTCAAACCAATAATGCTTCTCGGGGGTGGCCACGTACA 558
Dd	798	GCGCGGAGGAGATCTACTTCTCAAACCAATAATGCTTCTCGGGGGTGGCCACGTACA 857
QY	559	CCGACAAAGTTGTTCAAGTTCGGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
Dd	858	CCGACAAAGTTGTTCAAGTTCGGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 917
QY	619	ACGTGCCCCGTGTGGCCAGCTCTTTGCCGGAAGCTCTGTGGCTGTGTGGACAGAA 678
Dd	918	ACGTGCCCCGTGTGTGGCCAGCTCTTTGCCGGAAGCTCTGTGGCTGTGTGGACAGAA 977
QY	679	AGGGCTCTGGAAGCTACTTATCTACATTGCCAATACGCTACGCTACGTTAATGTGGGCCCTG 738
Dd	978	AGGGCTCTGGAAGCTACTTATCTACATTGCCAATACGCTACGCTACGTTAATGTGGGCCCTG 1037
QY	739	ATGCCCTCATTTGAATGGAACCTGAGCGCAGTACCTCTCCGGGGCATTTCTGCGCTCA 798
Dd	1038	ATGCCCTCATTTGAATGGAACCTGAGCGCAGTACCTCTCCGGGGCATTTCTGCGCTCA 1097
QY	799	GAGATGTGCTCTGAGGCTGGGTGAGCAATATACAGGGGCCGAGCGCTCAGCGTG 858
Dd	1098	GAGATGTGCTCTGAGGCTGGGTGAGCAATATACAGGGGCCGAGCGCTCAGCGTG 1157
QY	859	GCCCCATCTTCAGCAGCTGCTCGGATATCTTCTGGCACAATGAGATGGCGCTAACT 918
Dd	1158	GCCCCATCTTCAGCAGCTGCTCGGATATCTTCTGGCACAATGAGATGGCGCTAACT 1217
QY	919	TCTTTTCCAAACCGGGCGATGGCACTTTGTGACGCTGCGGCGAGTGGTGTGG 978
Dd	1218	TCTTTTCCAAACCGGGCGATGGCACTTTGTGACGCTGCGGCGAGTGGTGTGG 1277
QY	979	ACGACCCCAACCGATGGCGAGGTGTGCCCTTGCTGACTTCAACCGTGATGGCAAG 1038

RESULT 4  
AAH23100  
ID AAH23100 standard; DNA; 2263 BP  
XX  
XX AAH23100;  
XX

DT	17-SEP-2001 (first entry)	QY	379	ACCGGCCCCAGAGACGGCTGGTGAACATCGGGTGCATGAGCGCAGCTCACCTTACTACG	438
XX					
DE	Osteoarthritis tissue-derived nucleic acid sequence #30.	Db	513	ACCGGCCCCAGAGACGGCTGGTGAACATCGGGTGCATGAGCGCAGTAACCCCTACTACG	572
XX					
DE	Osteoarthritis tissue-derived nucleic acid sequence #30.	QY	439	CGCTGCCGAGCCGACAGGGGACGCCATGGGGTCAAGCTGCGACATCGACAGCGGACG	498
XX					
KW	Osteoarthritis; infectious disorder; Crohn's disease; sepsis; human;	Db	573	CGCTGCCGAGCCGACAGGGGACGCCATGGGGTCAAGCTGCGACATCGACAGCGGACG	632
KW	wound healing; osteopathic; anti-arthritis; anti-inflammatory; vulnerary;				
XX	antibacterial; antiallergic; ds.	QY	499	GCGGGAGGAGATCTACTTCTCAACCAATATAGCTTCTCGGGGTGCGCAGGTACA	558
OS	Homo sapiens.				
XX		Db	633	GCGGGAGGAGATCTACTTCTCAACCAATATAGCTTCTCGGGGTGCGCAGGTACA	692
XX					
PN	WO200153531-A2.	QY	559	CCGACAAAGTGTTCAGGTTCCGCAATACCGGTGGGAAGACATCTGAGCGATGAGGTCA	618
XX					
PD	26-JUL-2001.	Db	693	CCGACAAAGTGTTCAGGTTCCGCAATACCGGTGGGAAGACATCTGAGCGATGAGGTCA	752
XX					
PF	18-JAN-2001; 2001WO-US000016.	QY	619	ACGTGGCCGCTGGTGGCCAGCCCTTTTGGGACGCTCTGTGGCTGTGTGGACAGAA	678
XX					
PR	18-JAN-2000; 2000US-0176523P.	Db	753	ACGTGGCCGCTGGTGGCCAGCCCTTTTGGGACGCTCTGTGGCTGTGTGGACAGAA	812
XX					
PA	(PHAA ) PHARMACIA CORP.	QY	679	AGGCTCTGGACGGTACTCTATCTACATTCGCAATTCAGCTTACGGTATGTGGCCCTG	738
XX					
PI	Phippard D, Vasanthakamur G, Dotson S, Ma X;	Db	813	AGGCTCTGGACGGTACTCTATCTACATTCGCAATTCAGCTTACGGTATGTGGCCCTG	872
XX					
DR	WPI; 2001-451914/48.	QY	739	ATGCCCTCATTTGAATGACACCTGAGCCAGTACCTCTCCCGGGCATTTCTGGCGCTCA	798
XX					
XX	Substantially purified protein, polypeptide or their fragments, used to	Db	873	ATGCCCTCATTTGAATGACACCTGAGCCAGTACCTCTCCCGGGCATTTCTGGCGCTCA	932
PT	identify a biologically active compound or composition and treat				
PT	mammalian osteoarthritis.	QY	799	GAGATGTGCTGCTGAGGCTGGGTGAGCAATATACAGGGGCGGAGCGTCAGCGTGG	858
XX					
PS	Claim 1; Page 115-116; 14pp; English.	Db	933	GAGATGTGCTGCTGAGGCTGGGTGAGCAATATACAGGGGCGGAGCGTCAGCGTGG	992
XX					
CC	Sequences AAH23071-23152 represent nucleic acid sequences derived from	QY	859	GCCCATCTCTACGACGAGTGGCTCGGATATCTTCTGCGAATGAGAAATGGCCCTAACT	918
CC	osteoarthritis tissues. The sequences are useful as probes and for the				
CC	diagnosis or prognosis of mammalian osteoarthritis. The polynucleotides	Db	993	GCCCATCTCTACGACGAGTGGCTCGGATATCTTCTGCGAATGAGAAATGGCCCTAACT	1052
CC	and polypeptides of the invention are useful for generating diagnostic				
CC	reagents as targets for small molecule drug development, generation of	QY	919	TCCTTTTCCAAACCGGGGAGTGGACCTTTTGTGACCTGGGCGCTGCTGTGTGGTGG	978
CC	therapeutics, and cloning genes. Specific antibodies are used to generate				
CC	enzyme linked immunosorbent assays for detection of osteoarthritis. The	Db	1053	TCCTTTTCCAAACCGGGGAGTGGACCTTTTGTGACCTGGGCGCTGCTGTGTGGTGG	1112
CC	invented molecules can be used to treat osteoarthritis or to analyse the				
CC	disease-modifying activity of osteoarthritis drugs. Other disorders	QY	979	ACGACCCACACAGCATGGCGAGGTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1038
CC	treatable using the nucleic acid sequences include atopic, inflammatory				
CC	and infectious disorders e.g. Crohn's disease and sepsis, and wound	Db	1113	ACGACCCACACAGCATGGCGAGGTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1172
CC	healing				
XX		QY	1039	TGGACATCTCTATGGCAATGGAATGGCCCGCCACCGCTCTATCTGCAATGAGCACCC	1098
SQ	Sequence 2263 BP; 485 A; 667 C; 676 G; 428 T; 0 U; 7 Other;				
	Query Match 87.0%; Score 1893; DB 4; Length 2263;	Db	1173	TGGACATCTCTATGGCAATGGAATGGCCCGCCACCGCTCTATCTGCAATGAGCACCC	1232
	Best Local Similarity 99.6%; Pred. No. 0;				
	Matches 1918; Conservative 0; Mismatches 5; Indels 2; Gaps 2;	QY	1099	ATGGGAAGTTCGCTTCGGGACATCGCTCACCCCAAGTTCTCCATGCGCTCCCTCTGTC	1158
QY	19 CCGGGACCGGGCTGGGAGCAAGCAGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCG	Db	1233	ATGGGAAGTTCGCTTCGGGACATCGCTCACCCCAAGTTCTCCATGCGCTCCCTGTC	1292
Db	153 CCGGGACCGGGCTGGGAGCAAGCAGGCGGGCGCGCGCGCGCGCGCGCGCGCGCG				
QY	79 GCGCGTTCACCGCCCTAGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	QY	1159	GACCGGTATCACCGCGCTTTTGAATGACAGAGCTGGAGATCTTCTTCAACAACA	1218
Db	213 GCGCGTTCACCGCCCTAGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	Db	1293	GACCGGTATCACCGCGCTTTTGAATGACAGAGCTGGAGATCTTCTTCAACAACA	1352
QY	139 ACCCGGCGATGTCAGAGATTTACCGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG	QY	1219	TTGCTTACCGACATCTCTCAGCCAAACCGCTCTTCCGCGCTCATCCGCTAGAGACGAG	1278
Db	273 ACCCGGCGATGTCAGAGATTTACCGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG	Db	1353	TTGCTTACCGACATCTCTCAGCCAAACCGCTCTTCCGCGCTCATCCGCTAGAGACGAG	1412
QY	199 AGGGGTCCAGCGGCTGAACCATGTTCTACTGCGATCACCCTAGTTCGCTCCTG	QY	1279	ACCCCTCTATCGAGGAGCTCAATCCCGGCGACCGCTTGGAGCTTGGAGCGCGGCGACAG	1338
Db	333 AGGGGTCCAGCGGCTGAACCATGTTCTACTGCGATCACCCTAGTTCGCTCCTG	Db	1413	ACCCCTCTATCGAGGAGCTCAATCCCGGCGACCGCTTGGAGCTTGGAGCGCGGCGACAG	1472
QY	259 ACTATGACGTAATCCACACAGCTCAACTATGTTGGAGTGTGAGTGTGAGTGTGAG	QY	1339	GGGCTGTGTGACCGACTTCGACGGAGACGGGATGTGGACCTCATCTTGTCCATGAG	1398
Db	393 ACTATGACGTAATCCACACAGCTCAACTATGTTGGAGTGTGAGTGTGAGTGTGAG	Db	1473	GGGCTGTGTGACCGACTTCGACGGAGACGGGATGTGGACCTCATCTTGTCCATGAG	1532
QY	319 ATGGGAGCTTTGAGATCGTGGCGGGGTACAATGGACCCAACTGGTTCTGAAGTATG	QY	1399	AGTCCATGGCTACCGCTGTTCGGGGCAATCGAGCTTCAACAACTGCTGC	1458
Db	453 ATGGGAGCTTTGAGATCGTGGCGGGGTACAATGGACCCAACTGGTTCTGAAGTATG	Db	1533	AGTCCATGGCTACCGCTGTTCGGGGCAATCGAGGCTTCAACAACTGCTGC	1592
		QY	1459	TGCAGTGTGTGCC-ACGCACCCCGTTTGGGGCCCTTTGCCAGGGGAGCTAAGGTCTGTGTC	1517







QY	1691	CTCAGTCTGGAGATCTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCACT	1750
Db	1603	CTCAGTCTGGAGATCTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCACT	1662
QY	1751	GGAGTGTGGCAAGGATCTCCAGCAGGAAATGGCCATTCATGATGACACCAATGATG	1810
Db	1663	GGAGTGTGGCAAGGATCTCCAGCAGGAAATGGCCATTCATGATGACACCAATGATG	1722
QY	1811	CATCCAGTTCCTTCATTCCTGCTCGAGCAACCCCGTATGTCTAACACTATGGAAG	1870
Db	1723	CATCCAGTTCCTTCATTCCTGCTCGAGCAACCCCGTATGTCTAACACTATGGAAG	1782
QY	1871	CTACAGTGTGGCAAGGATCTCCAGCAGGAAATGGCCATTCATGATGACACCAATGATG	1930
Db	1783	CTACAGTGTGGCAAGGATCTCCAGCAGGAAATGGCCATTCATGATGACACCAATGATG	1842
QY	1931	AGCCTGCTGGG 1942	
Db	1843	AGCCTGCTGGG 1854	
RESULT 6			
AAD29791			
ID	AAD29791 standard; cDNA; 2039 BP.		
XX	AAD29791;		
AC	17-MAY-2002 (first entry)		
DT	Human SCIM-1 splice variant A cDNA.		
XX	Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;		
XX	extracellular matrix protein; tolerogenic protein; Grave's disease;		
KW	autoimmune disorder; juvenile arthritis; primary glomerulonephritis;		
KW	polyarthritis; osteoarthritis; Sjogren's disease; myasthenia gravis;		
KW	rheumatoid arthritis; Addison's disease; primary biliary sclerosis;		
KW	uveitis; systemic lupus erythematosus; inflammatory bowel disease;		
KW	multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.		
XX	Homo sapiens.		
XX	Key		
XX	Location/Qualifiers		
FT	59..1972		
FT	/*tag= a		
FT	/product= "Human SCIM-1 splice variant A protein"		
XX	WO200206478-A2.		
PN	24-JAN-2002.		
XX	09-JUL-2001; 2001WO-EP007888.		
XX	13-JUL-2000; 2000EP-00202495.		
XX	(ALKU ) AKZO NOVEL NV.		
XX	Heus HC, Nelissen RLH, Meeuwisse CML;		
XX	WPI; 2002-179796/23.		
XX	P-PSDB; AAE18680.		
XX	New extracellular matrix protein useful for prevention of inflammatory		
PT	diseases, more specifically in induction of T cell tolerance to the		
PT	protein in patients suffering from rheumatoid arthritis.		
XX	Claim 4; Page 35-36; 45pp; English.		
XX	The invention relates to human extracellular matrix protein i.e		
CC	Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic		
CC	acid molecules encoding such proteins. Sequences of the invention are		
CC	useful in a screening assay for the identification of tolerogenic		
CC	polypeptides. SCIM proteins are useful in therapy and for manufacture of		
CC	a pharmaceutical preparation against inflammatory diseases and for		

induction of immunological tolerance to an autoantigen in patients suffering from autoimmune disorders, more specifically rheumatoid arthritis. The inflammatory disorders treated include diseases like Grave's diseases, juvenile arthritis, primary glomerulonephritis, polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis, rheumatoid arthritis, Addison's disease, primary biliary sclerosis, uveitis, systemic lupus erythematosus, inflammatory bowel disease, multiple sclerosis and diabetes. Polypeptides of the invention are also useful in a diagnostic method for the detection of activated autoreactive T cells. The present sequence is human SCIM-1 splice variant A cDNA

Sequence 2039 BP; 448 A; 586 C; 596 G; 409 T; 0 U; 0 Other;

Query Match	83.0%;	Score 1807.8;	DB 6;	Length 2039;
Best Local Similarity	99.94;	Pred. No. 0;		
Matches 1809;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	131	GAGCGGTGACCCCGGATGTCAGGATGTTACCGTTCCTGCTGCTCTGTTCTGCC	190	
Db	43	GCGCGTGAACCCGGCATGTCAGGATGTTACCGTTCCTGCTGCTCTGTTCTGCC	102	
QY	191	CATCACTGAGGGTCCAGCGGCTGAACCCATGTTCACTGCACTCACTCACTTCT	250	
Db	103	CATCACTGAGGGTCCAGCGGCTGAACCCATGTTCACTGCACTCACTCACTTCT	162	
QY	251	GCCTCTGACTGACAGTAATCCACCCAGCTCAACTATGTTGGTGGCAGTTACTGAT	310	
Db	163	GCCTCTGACTGACAGTAATCCACCCAGCTCAACTATGTTGGTGGCAGTTACTGAT	222	
QY	311	GGACCATGATGGGACCTTTGAGATCGTGGCGGGTACAATGGACCCCAACCTGTTCT	370	
Db	223	GGACCATGATGGGACCTTTGAGATCGTGGCGGGTACAATGGACCCCAACCTGTTCT	282	
QY	371	GAAATGATGACCGGCGCAGAGCGGTGGTGAACATCGCGTCACTGAGCGCAGCTCACC	430	
Db	283	GAAATGATGACCGGCGCAGAGCGGTGGTGAACATCGCGTCACTGAGCGCAGCTCACC	342	
QY	431	CTACTACGCGCTCGGGACCGGACGGGACGCGCATTTGGGTCAAGCTTCGACATCGA	490	
Db	343	CTACTACGCGCTCGGGACCGGACGGGACGCGCATTTGGGTCAAGCTTCGACATCGA	402	
QY	491	CGGGACCGCGGAGAGATCTACTTCTCAACACCAATAATGCTTCTCGGGGTGGC	550	
Db	403	CGGGACCGCGGAGAGATCTACTTCTCAACACCAATAATGCTTCTCGGGGTGGC	462	
QY	551	CAGGTACACCGCAAGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCTGAGCGA	610	
Db	463	CAGGTACACCGCAAGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCTGAGCGA	522	
QY	611	TGAGGTCAACGTGGCCGCTGGTGGCCAGCTTCTTGGGACGCTCTGGCCCTGTGT	670	
Db	523	TGAGGTCAACGTGGCCGCTGGTGGCCAGCTTCTTGGGACGCTCTGGCCCTGTGT	582	
QY	671	GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCAATACGCTACGGTAATGT	730	
Db	583	GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCAATACGCTACGGTAATGT	642	
QY	731	GGGCCCTGATGCCCTCATTTGAATGAACCTGAGGCCAGTCACTTCCCGGGCATTTCT	790	
Db	643	GGGCCCTGATGCCCTCATTTGAATGAACCTGAGGCCAGTCACTTCCCGGGCATTTCT	702	
QY	791	GGCGCTCAGAGATGGCTGCTGAGCTGGGTGAGCAATATACAGGGGGCGAGGCT	850	
Db	703	GGCGCTCAGAGATGGCTGCTGAGCTGGGTGAGCAATATACAGGGGGCGAGGCT	762	
QY	851	CAGCGTGGGGCCCATCTCAGCAGAGTGCCTCGATATCTTCTGCAATAGAAATGG	910	
Db	763	CAGCGTGGGGCCCATCTCAGCAGAGTGCCTCGATATCTTCTGCAATAGAAATGG	822	
QY	911	GCCTAACTTCTTTTCCAAACCGGGGCGATGGCCTTTTGGACGCTCGGCCAGTGCC	970	
Db	823	GCCTAACTTCTTTTCCAAACCGGGGCGATGGCCTTTTGGACGCTCGGCCAGTGCC	882	

RESULT 7

AAD29793

ID AAD29793 standard; cDNA; 2145 BP.

XX

QY	971	TGTTGTGACGACCCGCCACAGCATGGCGAGGTGTGCCCTGGCTGACTTCAACCGTGA	1030	
Db	883	TGTTGTGACGACCCGCCACAGCATGGCGAGGTGTGCCCTGGCTGACTTCAACCGTGA	942	
QY	1031	TGGCAAGTGGACATCGTCTATGGCAACTGGAATGGCCGCCACCGCTCTATCTGCAAA	1090	
Db	943	TGGCAAGTGGACATCGTCTATGGCAACTGGAATGGCCGCCACCGCTCTATCTGCAAA	1002	
QY	1091	GAGCACCATGGGAAGTCCGCTTCCGGGACATCGCTCAACCCAAAGTTCCTCATGCCCTC	1150	
Db	1003	GAGCACCATGGGAAGTCCGCTTCCGGGACATCGCTCAACCCAAAGTTCCTCATGCCCTC	1062	
QY	1151	CCCTGTCGCAACCGTCTATCAACCGCGACTTTGCAATGACAGAGCTGGAGATCTTCTT	1210	
Db	1063	CCCTGTCGCAACCGTCTATCAACCGCGACTTTGCAATGACAGAGCTGGAGATCTTCTT	1122	
QY	1211	CAACAACATTTGCTACCGGAGCTTCTCAGCAACCGCTTCCGGGTCTATCGTAGAGA	1270	
Db	1123	CAACAACATTTGCTACCGGAGCTTCTCAGCAACCGCTTCCGGGTCTATCGTAGAGA	1182	
QY	1271	GCACGAGACCCCTCATCGAGGAGCTCAATCCCGCGAGCGCTTGGAGCTCGAGGCGG	1330	
Db	1183	GCACGAGACCCCTCATCGAGGAGCTCAATCCCGCGAGCGCTTGGAGCTCGAGGCGG	1242	
QY	1331	GGGCAACAGGGGTGTGGTGAACCGACTTCGACGAGACGGGATGCTGACCTCATCTTGT	1390	
Db	1243	GGGCAACAGGGGTGTGGTGAACCGACTTCGACGAGACGGGATGCTGACCTCATCTTGT	1302	
QY	1391	CCATGAGAGTCCATGGCTCAGCGCTGTCCTGCTTCCGGGGCAATCAGGGCTTCAACAA	1450	
Db	1303	CCATGAGAGTCCATGGCTCAGCGCTGTCCTGCTTCCGGGGCAATCAGGGCTTCAACAA	1362	
QY	1451	CAACTGGCTGCGAGTGGTGCACGACCCCGTTTGGGGCTTCCAGGGGAGCTAAGT	1510	
Db	1363	CAACTGGCTGCGAGTGGTGCACGACCCCGTTTGGGGCTTCCAGGGGAGCTAAGT	1422	
QY	1511	CGTGTCTTACACCAAGAGAGTGGGGCCCACTTGAGGATCATCGAGGGGGCTCAGGCTA	1570	
Db	1423	CGTGTCTTACACCAAGAGAGTGGGGCCCACTTGAGGATCATCGAGGGGGCTCAGGCTA	1482	
QY	1571	CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCTTGGGAGAGATGAAGCCAGAGTGT	1630	
Db	1483	CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCTTGGGAGAGATGAAGCCAGAGTGT	1542	
QY	1631	GGAGGTGACGTCGGCCAGATGCGCAAGATGCTGAGCCGGAACGTGGCCAGCGGAGATGA	1690	
Db	1543	GGAGGTGACGTCGGCCAGATGCGCAAGATGCTGAGCCGGAACGTGGCCAGCGGAGATGA	1602	
QY	1691	CTCAGTGTGGAGATCTCTTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT	1750	
Db	1603	CTCAGTGTGGAGATCTCTTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT	1662	
QY	1751	GGAGTGTGCGCAAGATTTCTCCAGCAGGAAAATGGCCATTTGAGGACCAATGAATG	1810	
Db	1663	GGAGTGTGCGCAAGATTTCTCCAGCAGGAAAATGGCCATTTGAGGACCAATGAATG	1722	
QY	1811	CATCAGTTCCTTCTGCTGTCCTCGAGCAAGCCCGTATGTGTCAACACCTATGGAAG	1870	
Db	1723	CATCAGTTCCTTCTGCTGTCCTCGAGCAAGCCCGTATGTGTCAACACCTATGGAAG	1782	
QY	1871	CTACAGGTGCGGACCAACAAAGAGTGCAGTGGGGCTACGAGCCCAACGAGGATGGCAC	1930	
Db	1783	CTACAGGTGCGGACCAACAAAGAGTGCAGTGGGGCTACGAGCCCAACGAGGATGGCAC	1842	
QY	1931	AGCTTGGCTGG 1941		
Db	1843	AGCTTGGCTGG 1853		

AC AAD29793;  
 XX 17-MAY-2002 (first entry)  
 XX Human SCIM-1 splice variant C cDNA.  
 XX Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;  
 KW extracellular matrix protein; tolerogenic protein; Grave's disease;  
 KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;  
 KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;  
 KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;  
 KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;  
 KW multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.  
 XX OS  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX CDS 59..1897  
 XX FT /\*tag= a  
 XX FT /product= "Human SCIM-1 splice variant C protein"  
 XX PN WO200206478-A2.  
 XX PD 24-JAN-2002.  
 XX PF 09-JUL-2001; 2001WO-EP007888.  
 XX PR 13-JUL-2000; 2000EP-00202495.  
 XX PA (ALKU) AKZO NOVEL NV.  
 XX PI Heus HC, Nelissen RLH, Meeuwisse CML;  
 XX WPI; 2002-179796/23.  
 XX DR P-PSDB; AAE18682.  
 XX PT New extracellular matrix protein useful for prevention of inflammatory  
 XX diseases, more specifically in induction of T cell tolerance to the  
 XX protein in patients suffering from rheumatoid arthritis.  
 XX Claim 4; Page 37; 45pp; English.  
 XX The invention relates to human extracellular matrix protein i.e  
 XX Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic  
 XX acid molecules encoding such proteins. Sequences of the invention are  
 XX useful in a screening assay for the identification of tolerogenic  
 XX polypeptides. SCIM proteins are useful in therapy and for manufacture of  
 XX a pharmaceutical preparation against inflammatory diseases and for  
 XX induction of immunological tolerance to an autoantigen in patients  
 XX suffering from autoimmune disorders, more specifically rheumatoid  
 XX arthritis. The inflammatory disorders treated include diseases like  
 XX Grave's diseases, juvenile arthritis, primary glomerulonephritis,  
 XX polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,  
 XX rheumatoid arthritis, Addison's disease, primary biliary sclerosis,  
 XX uveitis, systemic lupus erythematosus, inflammatory bowel disease,  
 XX multiple sclerosis and diabetes. Polypeptides of the invention are also  
 XX useful in a diagnostic method for the detection of activated autoreactive  
 XX T cells. The present sequence is human SCIM-1 splice variant C cDNA  
 XX Sequence 2145 BP; 497 A; 604 C; 611 G; 433 T; 0 U; 0 Other;  
 XX Query Match 83.0%; Score 1807.8; DB 6; Length 2145;  
 XX Best Local Similarity 99.9%; Pred. No. 0;  
 XX Matches 1809; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX 131 GAGCGGTGACCCCGGATGCCAGATGTTACCGTTCTCTGCTGCTCTGTTCTGCC 190  
 XX 43 GCGCGGTGACCCCGGATGCCAGATGTTACCGTTCTCTGCTGCTCTGTTCTGCC 102  
 XX 191 CATCACTGAGGGTCCAGGGGCTGAACCCATGTTTCATGAGTCAACCACTCACTTCT 250  
 XX 103 CATCACTGAGGGTCCAGGGGCTGAACCCATGTTTCATGAGTCAACCACTCACTTCT 162

QY 251 GCCTCTGACTATGACAGTAATCCACCAGCTCAACTATATGTTGGCAGTTACTGATGT 310  
 Db 163 GCCTCTGACTATGACAGTAATCCACCAGCTCAACTATATGTTGGCAGTTACTGATGT 222  
 QY 311 GGACCATGATGGGACTTTTGAGATCGTGTGGGGGTACAAATGGACCAACCTGGTTCT 370  
 Db 223 GGACCATGATGGGACTTTTGAGATCGTGTGGGGGTACAAATGGACCAACCTGGTTCT 282  
 QY 371 GAAGTATGACCGGCCAGAAAGCGCTGGTGAACAATCCGGTTCGATGAGCGAGCTCACC 430  
 Db 283 GAAGTATGACCGGCCAGAAAGCGCTGGTGAACAATCCGGTTCGATGAGCGAGCTCACC 342  
 QY 431 CTACTACCGCTGCGGGACCGGCGAGGGAACGCCAATGGGGTCAAGCTTGGCAATCGA 490  
 Db 343 CTACTACCGCTGCGGGACCGGCGAGGGAACGCCAATGGGGTCAAGCTTGGCAATCGA 402  
 QY 491 CGGGAGACGGCCGGGAGGAGATCTACTTCTCAACCAATATGCTTCTCGGGGTGGC 550  
 Db 403 CGGGAGACGGCCGGGAGGAGATCTACTTCTCAACCAATATGCTTCTCGGGGTGGC 462  
 QY 551 CACGTACACCGACAAAGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610  
 Db 463 CACGTACACCGACAAAGTTGTTCAAGTTCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522  
 QY 611 TGAGTCAACGTGGCCCGTGGTGGCAGGCTCTTTGCGGACGCTCTGTGGCCTGTGT 670  
 Db 523 TGAGTCAACGTGGCCCGTGGTGGCAGGCTCTTTGCGGACGCTCTGTGGCCTGTGT 582  
 QY 671 GGACAGAAAGGCTCTGACGCTACTCTATCTACATTCGCAATACGCTACGGAATGCT 730  
 Db 583 GGACAGAAAGGCTCTGACGCTACTCTATCTACATTCGCAATACGCTACGGAATGCT 642  
 QY 731 GGGCCCTGATGCCCTCATTTGAAATGGAACCTGAGGCCAGTGACCTCTCCCGGGGCTTCT 790  
 Db 643 GGGCCCTGATGCCCTCATTTGAAATGGAACCTGAGGCCAGTGACCTCTCCCGGGGCTTCT 702  
 QY 791 GGGCGCTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAATAATACAGGGGGCGAGGCGT 850  
 Db 703 GGGCGCTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAATAATACAGGGGGCGAGGCGT 762  
 QY 851 GAGCGTGGGCCCTCTCAGCAGCAGTGCCTCGGATATCTTCTCGCAATGAGATGG 910  
 Db 763 GAGCGTGGGCCCTCTCAGCAGCAGTGCCTCGGATATCTTCTCGCAATGAGATGG 822  
 QY 911 GCCTAACTTCTTTTCCAAACCGGGCGATGGCACTTTTGTGAGCGCTCGGCCAGTGC 970  
 Db 823 GCCTAACTTCTTTTCCAAACCGGGCGATGGCACTTTTGTGAGCGCTCGGCCAGTGC 882  
 QY 971 TGGTGTGACGACACCCCAACGATGGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTGA 1030  
 Db 883 TGGTGTGACGACACCCCAACGATGGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTGA 942  
 QY 1031 TGGCAAGTGGACATCGTCTATGGCACTGGAAATGGCCCCCACCCTCTATCTGCAAT 1090  
 Db 943 TGGCAAGTGGACATCGTCTATGGCACTGGAAATGGCCCCCACCCTCTATCTGCAAT 1002  
 QY 1091 GAGCACCATGGAAGTCCGCTTCGGGACATCGCTTCACCCCAAGTTCTCCATGCGCTC 1150  
 Db 1003 GAGCACCATGGAAGTCCGCTTCGGGACATCGCTTCACCCCAAGTTCTCCATGCGCTC 1062  
 QY 1151 CCCTGTCCGACGGTCTATCACCGCGACTTTTGACAAATGACGAGAGCTGAGATCTCTT 1210  
 Db 1063 CCCTGTCCGACGGTCTATCACCGCGACTTTTGACAAATGACGAGAGCTGAGATCTCTT 1122  
 QY 1211 CAACAACTGTCTACCGGAGCTCTCAGCAACCGCTCTTCCCGGTCTATCCGTAGAGA 1270  
 Db 1123 CAACAACTGTCTACCGGAGCTCTCAGCAACCGCTCTTCCCGGTCTATCCGTAGAGA 1182  
 QY 1271 GACCGGAGACCCCTCTCATCGAGAGCTCAATCCCGCGAGCGCTTGAGCGCTGAGGCGG 1330  
 Db 1183 GACCGGAGACCCCTCTCATCGAGAGCTCAATCCCGCGAGCGCTTGAGCGCTGAGGCGG 1242  
 QY 1331 GGGCACAGGGGGTGTGGTGACCGACTTTCGACGGAGAGCGGATGCTGGACCTCATCTTGTG 1390





```
QY 327 TTTGAGATCGTCTGGGGGTTAAATGGAACCAACTGTGTTCTGAAGTATGACCGGGCC 386
DB 181 TTTGAGATCGTCTGGGGGTTAAATGGAACCAACTGTGTTCTGAAGTATGACCGGGCC 240
QY 387 CAGAAGCGGCTGGTGAACATCGCGTGCATGAGCGCAGCTCACCTACTACGCGCTGCGG 446
DB 241 CAGNAGCGGCTGGTGAACATCGCGTGCATGAGCGCAGCTCACCTACTACGCGCTGCGG 300
QY 447 GACCGGAGGGAAACGCAATTTGGGGTCAACAGCCTGCGACATCGACGGGAGCGCCGGGAG 506
DB 301 GACCGGAGGGAAACGCAATCGGGTCAACAGCCTGCGACATCGACGGGAGCGCCGGGAG 360
QY 507 GAGATCTACTTCTCAACACCAATAATGCTCTCTCGGGGTGGCCAGCTACACGACAAG 566
DB 361 GAGATCTACTTCTCAACACCAATAATGCTCTCTCGGGGTGGCCAGCTACACGACAAG 420
QY 567 TTCTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAACGTGGCC 626
DB 421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAACGTGGCC 480
QY 627 CGTGGTGTGCCAGCCTCTTTGCGGACGCTCTGTGSCCTGTGTGGACAGAAAAGGCTCT 686
DB 481 CGTGGTGTGCCAGCCTCTTTGCGGACGCTCTGTGSCCTGTGTGGACAGAAAAGGCTCT 540
QY 687 GGNCGTACTCTATCTACATTTGCCAATTAGCCTTACGGTAAATGTGGCCCTGATGCCCTC 746
DB 541 GGAAGCTACTCTATCTACATTTGCCAATTAGCCTTACGGTAAATGTGGCCCTGATGCCCTC 600
QY 747 ATTGAATGAGCCTCAGGCGAGTACCTCTCCGGGGCAATCTCTGGCGCTCAGAGATGTG 806
DB 601 ATTGAATGAGCCTCAGGCGAGTACCTCTCCGGGGCAATCTCTGGCGCTCAGAGATGTG 660
QY 807 GTGTCTGAGGCTGGGTTCAGCAATAATACAGGGGGCCGAGCGCTCAGCGTGGGCCCAATC 866
DB 661 GCTGCTGAGGCTGGGTTCAGCAATAATACAGGGGGCCGAGCGCTCAGCGTGGGCCCAATC 720
QY 867 CTCAGCAGAGTGCCTTCGATATCTTCTGGCAATGAGAGTGGCCCTAACTTCTTTTC 926
DB 721 CTCAGCAGAGTGCCTTCGATATCTTCTGGCAATGAGAGTGGCCCTAACTTCTTTTC 780
QY 927 CACAAACGGGGGATGGCACTTTGTGGACGCTGCGGCCAGTCTGCTGTGGACGACCC 986
DB 781 CACAAACGGGGGATGGCACTTTGTGGACGCTGCGGCCAGTCTGCTGTGGACGACCC 840
QY 987 CACCAAGATGGGCGAGGTGTGGCCCTGGTGAATTCACCGTGTATGCAAAAGTGGACATC 1046
DB 841 CACCAAGATGGGCGAGGTGTGGCCCTGGTGAATTCACCGTGTATGCAAAAGTGGACATC 900
QY 1047 GTCTATGGCACTGGGAATGGCCCCCAGCCGCTCTATCTGCAATGAGACCCATGGGAAG 1106
DB 901 GTCTATGGCACTGGGAATGGCCCCCAGCCGCTCTATCTGCAATGAGACCCATGGGAAG 960
QY 1107 GTCCGCTTCGGGACATCGCTCACCAAGTTCTCCATGCCCTCCCTGTCGCCACGGTC 1166
DB 961 GTCCGCTTCGGGACATCGCTCACCAAGTTCTCCATGCCCTCCCTGTCGCCACGGTC 1020
QY 1167 ATCACCGCGACTTTGACAATGACAGGAGCTGGAGATCTTCTTCAACAACATTTGCTTAC 1226
DB 1021 ATCACCGCGACTTTGACAATGACAGGAGCTGGAGATCTTCTTCAACAACATTTGCTTAC 1080
QY 1227 CGCAGTCTCTCAGCCAAACCGCTCTTCGCGGTATCGGTAGAGAGCAGGAGACCCCTC 1286
DB 1081 CGCAGTCTCTCAGCCAAACCGCTCTTCGCGGTATCGGTAGAGAGCAGGAGACCCCTC 1140
QY 1287 ATCAGAGCTCAATCCCGCGCAGCCTTGGAGCCTGAGCGCGGGGCAAGGGGTGTG 1346
DB 1141 ATCAGAGCTCAATCCCGCGCAGCCTTGGAGCCTTGGAGCGCGGGGCAAGGGGTGTG 1200
QY 1347 GTGACCGATTTCAGCGGAGACGGGATGCTGGACCTCATCTTGTCCATGGAGAGTCCATG 1406
DB 1201 GTGACCGATTTCAGCGGAGACGGGATGCTGGACCTCATCTTGTCCATGGAGAGTCCATG 1260
```

```
QY 1407 GCTCAGCGGTGTCGTCTTCCGGGGCAATCAGGGCTTCAACAACAATCGCTGCGAGTG 1466
DB 1261 GCTCAGCGGTGTCGTCTTCCGGGGCAATCAGGGCTTCAACAACAATCGCTGCGAGTG 1320
QY 1467 GTCCCAACGACCCGGTTTGGGGCTTTTGCAGGGGAGCTAAGGTCTGTCTTACACCAAG 1526
DB 1321 GTGCCACGACCCGGTTTGGGGCTTTTGCAGGGGAGCTAAGGTCTGTCTTACACCAAG 1380
QY 1527 AAGAGTGGGCCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
DB 1381 AAGAGTGGGCCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440
QY 1587 CCCGTGGCACACTTTGGCCTGGGGAAGATGAAGCAGCAGTGTGGAGTGAACGTGGCCA 1646
DB 1441 CCCGTGGCACACTTTGGCCTGGGGAAGATGAAGCAGCAGTGTGGAGTGAACGTGGCCA 1500
QY 1647 GATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGAGATGAACCTCAGTCTGGAGATC 1706
DB 1501 GATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGAGATGAACCTCAGTCTGGAGATC 1560
QY 1707 CTCTACCCCGGATGAGGACACACTTTCAGGACCCAGCCCCCACTGGAGTGTGCCAAGGA 1766
DB 1561 CTCTACCCCGGATGAGGACACACTTTCAGGACCCAGCCCCCACTGGAGTGTGCCAAGGA 1620
QY 1767 TTCTCCAGCAGGAAATGGCCATTTGCATGGACACCAATGAATGCCATTCACATTC 1826
DB 1621 TTCTCCAGCAGGAAATGGCCATTTGCATGGACACCAATGAATGCCATTCACATTC 1680
QY 1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGTTCGGGACC 1886
DB 1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGTTCGGGACC 1740
QY 1887 AACAGAAGTGCAGTCCGGGCTACGAGCCCCAACAGAGATGGCACAGCTTCGTGGG 1942
DB 1741 AACAGAAGTGCAGTCCGGGCTACGAGCCCCAACAGAGATGGCACAGCTTCGTGGG 1796
```

RESULT 10  
ACC50982

ID ACC50982 standard; cdna; 2507 BP.

XX ACC50982;

XX 12-JUN-2003 (first entry)

XX Human bladder cancer associated cdna sequence SEQ ID NO:57.

XX Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.

XX Homo sapiens.

XX WO2003003906-A2.

XX 16-JAN-2003.

XX 03-JUL-2002; 2002WO-US021338.

XX 03-JUL-2001; 2001US-0302814P.

XX 03-AUG-2001; 2001US-0310099P.

XX 08-NOV-2001; 2001US-0343705P.

XX 13-NOV-2001; 2001US-0350666P.

XX 12-APR-2002; 2002US-0372246P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Mack DH, Aziz N;

XX WPI; 2003-201532/19.

XX P-PSDB; ABR48171.

XX Detecting a bladder cancer-associated transcript in a cell from a patient, comprises contacting a biological sample from the patient with a bladder cancer-associated polynucleotide or antibody.













61	AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTGTGGCGGGTACAAATGACC	120
359	CAACCTGGTCTTGAAGTATGACCGGGCCAGAAAGCGGTGTGTGAACATCGCGGTGCAATGA	418
121	CAACCTGGTCTTGAAGTATGACCGGGCCAGAAAGCGGTGTGTGAACATCGCGGTGCAATGA	180
419	GGCAGCTCACCTACTACCGCTCGGACCGGACGGGAAACGCCATTGGGGTCAACAG	478
181	GGCAGCTCACCTACTACCGCTCGGACCGGACGGGAAACGCCATTGGGGTCAACAG	240
479	CTGCGACATCGACGGGACCGCGGAGAGATCTACTTCTCTCAACCAATAATAGCTT	538
241	CTGCGACATCGACGGGACCGCGGAGAGATCTACTTCTCTCAACCAATAATAGCTT	300
539	CTGGGGTGGCCACGTACACGACAAAGTGTTCAGTTTCGCAATAACCGTGGGAAGA	598
301	CTGGGGTGGCCACGTACACGACAAAGTGTTCAGTTTCGCAATAACCGTGGGAAGA	360
599	CATCCTGAGCGATGAGGTCAACGTGGCCCGTGTGTGGCCAGCTCTTTGCGGACGCTC	658
361	CATCCTGAGCGATGAGGTCAACGTGGCCCGTGTGTGGCCAGCTCTTTGCGGACGCTC	420
659	TGTGGCCTGTGTGGACAGAAAGGCTCTGGAACCTACTCTACTACTGATTCGCAATTCGC	718
421	TGTGGCCTGTGTGGACAGAAAGGCTCTGGAACCTACTCTACTACTGATTCGCAATTCGC	480
719	CTACGGTAATGTGGGCCCTGATCGCTCATTTGAAATGGACCTGAGGCCAGTGAACCTCTC	778
481	CTACGGTAATGTGGGCCCTGATCGCTCATTTGAAATGGACCTGAGGCCAGTGAACCTCTC	540
779	CCGGGCGAATCTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAATATACAGG	838
541	CCGGGCGAATCTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAATATACAGG	600
839	GGGCGGAGGCTCAGCGTGGGCCCTCATCTCAGCAGAGTGCTCGGATCTCTTCGGA	898
601	GGGCGGAGGCTCAGCGTGGGCCCTCATCTCAGCAGAGTGCTCGGATCTCTTCGGA	660
899	CAATGAGAAATGGGCTTAACCTCTTTTCCAAACCGGGGCGATGGACCTTTGTGGACGC	958
661	CAATGAGAAATGGGCTTAACCTCTTTTCCAAACCGGGGCGATGGACCTTTGTGGACGC	720
959	TGGGCCAGTCTGTGTGGACGACCCCGCCAGCATGGCGAGGTGTGCGCCTGCTCGA	1018
721	TGGGCCAGTCTGTGTGGACGACCCCGCCAGCATGGCGAGGTGTGCGCCTGCTCGA	780
1019	CTTCAACCGTGAATGCAAGTGGACATGCTATGGCAATGGAAATGGCCCCCAGCGCT	1078
781	CTTCAACCGTGAATGCAAGTGGACATGCTATGGCAATGGAAATGGCCCCCAGCGCT	840
1079	CTATCTGAAATGAGCACCCATGGGAGTCCGCTTCGGGACATCGCTCACCCAAAGTT	1138
841	CTATCTGAAATGAGCACCCATGGGAGTCCGCTTCGGGACATCGCTCACCCAAAGTT	900
1139	CTCCATGCCCCTCCCTGTCCGACGGTCAATCACCGCCAGCTTTGACAAATGACGAGACT	1198
901	CTCCATGCCCCTCCCTGTCCGACGGTCAATCACCGCCAGCTTTGACAAATGACGAGACT	960
1199	GGAGATCTTCTTCAACAACATTTGCTACCGAGTCTCTCAGCAACCGCTCTTCGCGCT	1258
961	GGAGATCTTCTTCAACAACATTTGCTACCGAGTCTCTCAGCAACCGCTCTTCGCGCT	1020
1259	CATCTGTAGAGACGAGAGACCCCTCATCGAGAGCTCAATCCCGGCGACGCTTTGGA	1318
1021	CATCTGTAGAGACGAGAGACCCCTCATCGAGAGCTCAATCCCGGCGACGCTTTGGA	1080
1319	GCCTGAGGCGCGGGCACAGGGGTTGTGGTACCGCATTTGCAACGAGACGGGATGCTTGA	1378
1081	GCCTGAGGCGCGGGCACAGGGGTTGTGGTACCGCATTTGCAACGAGACGGGATGCTTGA	1140
1379	CCTCATCTTGTCCATGGAGAGTCCATGGCTCAGCGCTGTCCGCTCTTCGGGGCAATCA	1438
1141	CCTCATCTTGTCCATGGAGAGTCCATGGCTCAGCGCTGTCCGCTCTTCGGGGCAATCA	1200

QY	1439	GGGCTTCAACAACACTGCTGGAGTGGTGCACGACCCCGCTTTGGGGCTTTGGCCAG	1499
DB	1201	GGGCTTCAACAACAACACTGCTGGAGTGGTGCACGACCCCGCTTTGGGGCTTTGGCCAG	1260
QY	1499	GGGAGCTAAGGTCGTCTCTACACCAAGAAGAGTGGGGCCCACTTGAGGATCATCGACGG	1558
DB	1261	GGGAGCTAAGGTCGTCTCTACACCAAGAAGAGTGGGGCCCACTTGAGGATCATCGACGG	1320
QY	1559	GGGCTCAGGCTACCTGTGTGAGATGAGAGCCGCTGGGCACACTTTGGCTCGGGGAGAGATGA	1618
DB	1321	GGGCTCAGGCTACCTGTGTGAGATGAGAGCCGCTGGGCACACTTTGGCTCGGGGAGAGATGA	1380
QY	1619	AGCCAGCAGTGTGGAGTGCAGTGGCCAGATGCCAAGATGGTGAGCGCGGAAACGTGCCACG	1678
DB	1381	AGCCAGCAGTGTGGAGTGCAGTGGCCAGATGCCAAGATGGTGAGCGCGGAAACGTGCCACG	1440
QY	1679	CGGGAGATGAATCTCAGTGTGGAGATCTCTACCCCCGGGATGAGGACACACTTCAGGA	1738
DB	1441	CGGGAGATGAATCTCAGTGTGGAGATCTCTACCCCCGGGATGAGGACACACTTCAGGA	1500
QY	1739	CCGAGCCCCACTGGAGTGTGGCCAAAGATTCTCCGACGAGAAAATGGCCATTCGATGGA	1798
DB	1501	CCGAGCCCCACTGGAGTGTGGCCAAAGATTCTCCGACGAGAAAATGGCCATTCGATGGA	1560
QY	1799	CACCAATGAATGCATCCAGTTCGATTCGTTGTCGCCCTCGAGACAAGCCCGTATGTGTCAA	1858
DB	1561	CACCAATGAATGCATCCAGTTCGATTCGTTGTCGCCCTCGAGACAAGCCCGTATGTGTCAA	1620
QY	1859	CACCTATGGAGCTACAGGTCGCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGGCCAA	1918
DB	1621	CACCTATGGAGCTACAGGTCGCGGACCAACAAGAAGTGCAGTCGAGTCGGGGCTACGAGGCCAA	1679
QY	1919	CGAGGATGGCACAGCCCTGGCTGG	1941
DB	1680	CGAGGATGGCACAGCCCTGGCTGG	1702

RESULT 14

```

Result 2
AAH14000
ID AAH14000 standard; cDNA; 2178 BP.
XX
XX AAH14000;
XX
XX AC
XX
XX AC
XX
DT 26-JUN-2001 (first entry)
XX
XX DE
XX DE Human cDNA sequence SEQ ID NO:11082.
XX
XX KW
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX OS
XX OS Homo sapiens.
XX
XX FN
XX FN EP1074617-A2.
XX
XX PD
XX PD 07-FEB-2001.
XX
XX PF
XX PF 28-JUL-2000; 2000EP-00116126.
XX
XX PR
XX PR 29-JUL-1999; 99JP-00248036.
XX
XX PR 27-AUG-1999; 99JP-00300253.
XX
XX PR 11-JAN-2000; 2000JP-00118776.
XX
XX PR 02-MAY-2000; 2000JP-00183767.
XX
XX PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX PA
XX PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX PI
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX DR
XX DR WPI; 2001-318749/34.
XX
XX PT
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX
XX PT length cDNAs defined in the specification, and for the detection and/or
XX

```

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.

PS Claim 8; SEQ ID NO 11082; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-  
XX length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification; where the  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
XX present invention

SQ Sequence 2178 BP; 477 A; 618 C; 655 G; 428 T; 0 U; 0 Other;

Query Match 64.7%; Score 1408.8; DB 4; Length 2178;

Best Local Similarity 97.0%; Pred. No. 0;

Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 435 TACGCGCTGCGGACCGGACCGGAGAAACCCATTTGGGTCTACAGCTCGCATCGACGGG 494  
DB 1 TACGCGCTGCGGACCGGACCGGAGAAACCCATCGGGGTCTACGCTCGCATCGACGGG 60  
QY 495 GACGCGCGGAGAGATCTACTTCTCTCAACACCAATAATGCTTCTCGGGGTGGCCACG 554  
DB 61 GACGCGCGGAGAGATCTACTTCTCTCAACACCAATAATGCTTCTCGGGGTGGCCACG 120  
QY 555 TACACGCGAAGTGTTCAGTTTCGCAATAACCGGTGGGAACATCTTCGAGGATGAG 614  
DB 121 TACACGCGAAGTGTTCAGTTTCGCAATAACCGGTGGGAACATCTTCGAGGATGAG 180  
QY 615 GTCAACGTGCGCGCTGTGTGGCCAGCGCTCTTTGCGGACGCTCTGTGGCTGTGTGGAC 674  
DB 181 GTCAACGTGCGCGCTGTGTGGCCAGCGCTCTTTGCGGACGCTCTGTGGCTGTGTGGAC 240  
QY 675 AGAAGGGCTCTGGAAGCTACTTATCTACTATTCATTTGCCAATTAAGCCCTACGGTATGTGGC 734  
DB 241 AGAAGGGCTCTGGAAGCTACTTATCTACTATTCATTTGCCAATTAAGCCCTACGGTATGTGGC 300  
QY 735 CCTGATGCGCTCTTCAATGAGCCCTGAGCCAGTGAACCTCTCCGCGGCAATTCGGCG 794  
DB 301 CCTGATGCGCTCTTCAATGAGCCCTGAGCCAGTGAACCTCTCCGCGGCAATTCGGCG 360  
QY 795 CTCAGAGATGTGGCTGCTGAGGTGGGTGAGCAATATACAGGGGCGGAGCGCTCAGC 854  
DB 361 CTCAGAGATGTGGCTGCTGAGGTGGGTGAGCAATATACAGGGGCGGAGCGCTCAGC 420  
QY 855 GTGGGCGCCCATCTCAGCAGCAGTGTCTGGATATCTTCTCGCAATAGAGATGTGGCCCT 914  
DB 421 GTGGGCGCCCATCTCAGCAGCAGTGTCTGGATATCTTCTCGCAATAGAGATGTGGCCCT 480  
QY 915 AACTTCTCTTTCACAAACCGGGCGATGGACCTTTGTGGAAGCTGCGGCGAGTGTGGT 974  
DB 481 AACTTCTCTTTCACAAACCGGGCGATGGACCTTTGTGGAAGCTGCGGCGAGTGTGGT 540  
QY 975 GTGGAGCAGCCCCCACCAGCATGGGCGAGGTGTGCGCCCTGCGTGTGACCTTCAACCGTATGCG 1034  
DB 541 GTGGAGCAGCCCCCACCAGCATGGGCGAGGTGTGCGCCCTGCGTGTGACCTTCAACCGTATGCG 600

QY 1035 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCAGCTCTATCTGCAATGAGC 1094  
DB 601 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCAGCTCTATCTGCAATGAGC 660  
QY 1095 ACCCATGGGAAGGTCCGCTTCCGGGACATCCGCTCAACCAAGTTCTTCATGCCCTCCCT 1154  
DB 661 ACCCATGGGAAGGTCCGCTTCCGGGACATCCGCTCAACCAAGTTCTTCATGCCCTCCCT 720  
QY 1155 GTCCGACGGTCTACACCGCGGACTTTGACAAATACCAAGGAGCTGGAGATCTTCTCAAC 1214  
DB 721 GTCCGACGGTCTACACCGCGGACTTTGACAAATACCAAGGAGCTGGAGATCTTCTCAAC 780  
QY 1215 AACATTGCTACCGCAGCTCTCTAGCCAAACCGCTCTTCCCGCTCATCCGTAGAGAGCAC 1274  
DB 781 AACATTGCTACCGCAGCTCTCTAGCCAAACCGCTCTTCCCGCTCATCCGTAGAGAGCAC 840  
QY 1275 GGAGACCCCTCTATCGAGGAGCTCAATCCCGGACGCTTTGGAGCTTGAGGGCCGGGGC 1334  
DB 841 GGAGACCCCTCTATCGAGGAGCTCAATCCCGGACGCTTTGGAGCTTGAGGGCCGGGGC 900  
QY 1335 ACAGGGGGTGTGTGACCGGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCAT 1394  
DB 901 ACAGGGGGTGTGTGACCGGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCAT 960  
QY 1395 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGCTTCAACAAACAC 1454  
DB 961 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGCTTCAACAAACAC 1020  
QY 1455 TGGCTGGAGTGTGCCACGACCCCGCTTTGGGGCTTTGCCAGGGGAGCTAAGGTCTGTG 1514  
DB 1021 TGGCTGGAGTGTGCCACGACCCCGCTTTGGGGCTTTGCCAGGGGAGCTAAGGTCTGTG 1080  
QY 1515 CTCTACACCAAGAGAGTGGGGCCCACTCAGGATCATCCAGCGGGGCTCAGGCTACTGT 1574  
DB 1081 CTCTACACCAAGAGAGTGGGGCCCACTCAGGATCATCCAGCGGGGCTCAGGCTACTGT 1140  
QY 1575 TGTGAGATGAGCCCGTGGGCACACTTTGGGCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1634  
DB 1141 TGTGAGATGAGCCCGTGGGCACACTTTGGGCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1200  
QY 1635 GTGACGTGGCCAGATGGCAAGATGTTGAGCCGGAACGTGGCCAGCGGGGAGATGAATCA 1694  
DB 1201 GTGACGTGGCCAGATGGCAAGATGTTGAGCCGGAACGTGGCCAGCGGGGAGATGAATCA 1260  
QY 1695 GTGCTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAG 1754  
DB 1261 GTGCTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAG 1319  
QY 1755 TGTGGCCAAGGATTTCTCCAGCAGGAATAATGGCCATTGCAATGACACCAATGAATGCATC 1814  
DB 1320 -----GACACCAATGAATGCATC 1337  
QY 1815 CAGTTCCTCATCTGTGTGCCCTCGAGCAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1874  
DB 1338 CAGTTCCTCATCTGTGTGCCCTCGAGCAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1397  
QY 1875 AGTGGCCGACCAACCAAGAGTGCAGTTCGGGGCTACGAGCCCAACGAGGATGGCAGACC 1934  
DB 1398 AGTGGCCGACCAACCAAGAGTGCAGTTCGGGGCTACGAGCCCAACGAGGATGGCAGACC 1457  
QY 1935 TGGCTGGG 1942  
DB 1458 TGGCTGGG 1465

RESULT 15

ACC50981

ID ACC50981 standard; cDNA; 2178 BP.

XX ACC50981;

AC ACC50981;

XX 12-JUN-2003 (first entry)

XX Human bladder cancer associated cDNA sequence SEQ ID NO:55.  
XX Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.  
XX Homo sapiens.  
XX WO2003003906-A2.  
XX 16-JAN-2003.  
XX 03-JUL-2002; 2002WO-US021338.  
XX 03-JUL-2001; 2001US-0302814P.  
XX 03-AUG-2001; 2001US-0310099P.  
XX 08-NOV-2001; 2001US-0343705P.  
XX 13-NOV-2001; 2001US-0350666P.  
XX 12-APR-2002; 2002US-0372246P.  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
XX Mack DH, Aziz N;  
XX WPI: 2003-201532/19.  
XX P-PSDB; ABR48170.  
XX Detecting a bladder cancer-associated transcript in a cell from a  
XX patient, comprises contacting a biological sample from the patient with a  
XX bladder cancer-associated polynucleotide or antibody.  
XX Claim 6; Page 247-248; 307pp; English.  
XX The present invention describes a method for detecting a bladder cancer-  
XX associated transcript in a cell from a patient. The method comprises  
XX contacting a biological sample from the patient with a polynucleotide  
XX that selectively hybridizes to a sequence that is 80 % identical to a  
XX table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059  
XX encode the human bladder cancer-associated proteins given in ABR48146 to  
XX ABR48242). Bladder cancer-associated sequences from the present invention  
XX have cytostatic activities, and can be used in antisense gene therapy and  
XX in vaccine production. The method can be used for detecting a bladder  
XX cancer-associated transcript in a cell from a patient. The method is  
XX useful in diagnosing or treating bladder cancer and in screening for  
XX compounds that modulate bladder cancer, such as hormones or antibodies.  
XX The nucleic acid molecules from the present invention may be used in  
XX various screening and diagnostic methods, and for gene therapy, vaccine  
XX and/or antisense/inhibition applications  
XX  
SQ Sequence 2178 BP; 477 A; 618 C; 655 G; 428 T; 0 U; 0 Other;  
Query Match 64.7%; Score 1408.8; DB 7; Length 2178;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;  
QY 435 TAGCGCTGGGACCGGAGCGGAGCGGATGGGTCACAGCCTGGGACATCGACGG 494  
DB 1 TAGCGCTGGGACCGGAGCGGAGCGGATGGGTCACAGCCTGGGACATCGACGG 60  
QY 495 GACGGCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCCTCTCGGGGGTGGCCACG 554  
DB 61 GACGGCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCCTCTCGGGGGTGGCCACG 120  
QY 555 TACACCGACAAAGTGTTCAGTTCGGAATAACCGGTGGGAAGACATCTCTGAGCGATGAG 614  
DB 121 TACACCGACAAAGTGTTCAGTTCGGAATAACCGGTGGGAAGACATCTCTGAGCGATGAG 180  
QY 615 GTCAACGTGGCGGTGGTGGCGAGCCTCTTTGCCGAGCGCTCTGTGGCCTGTGTGGAC 674  
DB 181 GTCAACGTGGCGGTGGTGGCGAGCCTCTTTGCCGAGCGCTCTGTGGCCTGTGTGGAC 240  
QY 675 AGAAGGGCTCTGGACGCTACTCTATCTACATTTGCCAATTCAGCCTACGTAATGTGGG 734  
DB 241 AGAAGGGCTCTGGACGCTACTCTATCTACATTTGCCAATTCAGCCTACGTAATGTGGG 300

QY 735 CCTGATGCCCTCATTGAAATGAGCCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCG 794  
DB 301 CCTGATGCCCTCATTGAAATGAGCCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCG 360  
QY 795 CTCAGAGATGTGGTGTGCTGAGCGTGGGTGAGCAAAATATACAGGGGGCCGAGGCGTCAGC 854  
DB 361 CTCAGAGATGTGGTGTGCTGAGCGTGGGTGAGCAAAATATACAGGGGGCCGAGGCGTCAGC 420  
QY 855 GTGGGCCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAAATGAGAAATGGGCT 914  
DB 421 GTGGGCCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAAATGAGAAATGGGCT 480  
QY 915 AACTTCTTTTCCACAAACCGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGTGTGT 974  
DB 481 AACTTCTTTTCCACAAACCGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGTGTGT 540  
QY 975 GTGGACGACCCCAACAGCAGTGGGCGAGGTGTGCGCTGTGACTTCAACCGTGTATGCG 1034  
DB 541 GTGGACGACCCCAACAGCAGTGGGCGAGGTGTGCGCTGTGACTTCAACCGTGTATGCG 600  
QY 1035 AAAAGTGGACATCTCTATGGCAACTGGAAATGSCCCCAACCGCTCTATCTGCAAAATGAGC 1094  
DB 601 AAAAGTGGACATCTCTATGGCAACTGGAAATGSCCCCAACCGCTCTATCTGCAAAATGAGC 660  
QY 1095 ACCCATGGGAAGTCCGCTTCCGGGACATCGCTCACCCAAAGTTCTCCATGCCCCTCCCT 1154  
DB 661 ACCCATGGGAAGTCCGCTTCCGGGACATCGCTCACCCAAAGTTCTCCATGCCCCTCCCT 720  
QY 1155 GTCCGACCGTCAATCACCCGCGACTTTGACAAATGACACGAGAGCTGTGAGATCTTCTCAAC 1214  
DB 721 GTCCGACCGTCAATCACCCGCGACTTTGACAAATGACACGAGAGCTGTGAGATCTTCTCAAC 780  
QY 1215 AACATTGCTACGACAGTCTCTCAGCAACCGCTCTCCGGTCAATCGTAGAGAGCAC 1274  
DB 781 AACATTGCTACGACAGTCTCTCAGCAACCGCTCTCCGGTCAATCGTAGAGAGCAC 840  
QY 1275 GGAGACCCCTCATCGAGAGCTCAATCCCGCGCGAGCGCTTGAGAGCTGAGGCCCGGGG 1334  
DB 841 GGAGACCCCTCATCGAGAGCTCAATCCCGCGCGAGCGCTTGAGAGCTGAGGCCCGGGG 900  
QY 1335 ACAGGGGGTGTGTGTGACCGCTTGCAGGAGAGCGGATGTGAGACCTCATCTTGTCCCAT 1394  
DB 901 ACAGGGGGTGTGTGTGACCGCTTGCAGGAGAGCGGATGTGAGACCTCATCTTGTCCCAT 960  
QY 1395 GGAGAGTCCATGCTCAGCGCTGTCCGCTTCCGGGGCAATCAGGCTTCAACAAACAC 1454  
DB 961 GGAGAGTCCATGCTCAGCGCTGTCCGCTTCCGGGGCAATCAGGCTTCAACAAACAC 1020  
QY 1455 TGCTCGGAGTGTGTGCGACCGCTTGGGGCTTTGGGGCTTGGCCAGGGAGCTAAGGTGCTG 1514  
DB 1021 TGCTCGGAGTGTGTGCGACCGCTTGGGGCTTTGGGGCTTGGCCAGGGAGCTAAGGTGCTG 1080  
QY 1515 CTCTACACCAAGAGTGGGGCCCACTGAGAGATCATCGACGGGGCTCAGGCTACCTG 1574  
DB 1081 CTCTACACCAAGAGTGGGGCCCACTGAGAGATCATCGACGGGGCTCAGGCTACCTG 1140  
QY 1575 TGTGAGATGAGTCCCTGCGACACTTTGGCTGGGGAAGATGAAGCAGCAGTGTGGAG 1634  
DB 1141 TGTGAGATGAGTCCCTGCGACACTTTGGCTGGGGAAGATGAAGCAGCAGTGTGGAG 1200  
QY 1635 GTGACGTGGCCATGGCAAGATGTGAGCCCGGAAAGTGGCCAGCGGGAGATGAACCTCA 1694  
DB 1201 GTGACGTGGCCATGGCAAGATGTGAGCCCGGAAAGTGGCCAGCGGGAGATGAACCTCA 1260  
QY 1695 GTGCTGAGATCTCTTACCCCGGGATGAGGACACACTTCAGGACCCGACCTGAG 1754  
DB 1261 GTGCTGAGATCTCTTACCCCGGGATGAGGACACACTTCAGGACCCGACCTGAG - 1319  
QY 1755 TGTGGCCAAAGATTCTCCACAGGAAATGGCCATTGTCATGAGACCAATGAATGATC 1814  
DB 1320 -----GACACCAATGAATGATC 1337

Tue Aug 10 12:45:35 2004

QY	1815	CAGTCCCATTCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACCTATGGAAGCTAC	1874
Db	1338	CAGTCCCATTCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACCTATGGAAGCTAC	1397
QY	1875	AGGTGCCGGACCAACAAGAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAGCC	1934
Db	1398	AGGTGCCGGACCAACAAGAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAGCC	1457
QY	1935	TGCGTGGG	1942
Db	1458	TGCGTGGG	1465

Search completed: August 9, 2004, 17:31:02  
Job time : 840 secs



mis Page Blank (uspto)

mis Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 17:17:00 ; Search time 179 Seconds  
(without alignments)

6749.319 Million cell updates/sec

Title: US-09-914-958B-35

Perfect score: 2177

Sequence: 1 cgggaggtcgcagccagcc.....tcctttgaaaaaaaaa 2177

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
  - 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
  - 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*
  - 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1345.8	61.8	1501	4	US-09-620-312D-1032	Sequence 1032, Ap
2	788	36.2	789	4	US-09-023-655-484	Sequence 484, App
3	109	5.0	707	4	US-09-023-655-209	Sequence 209, App
C 4	58.2	2.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 5	54.8	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
6	47.8	2.2	1221	4	US-09-252-991A-10929	Sequence 10929, A
7	47.8	2.2	1269	4	US-09-252-991A-11061	Sequence 11061, A
C 8	47.8	2.2	1641	4	US-09-252-991A-11134	Sequence 11134, A
9	46.8	2.1	501	4	US-09-252-991A-10991	Sequence 10991, A
10	43.8	2.0	436	4	US-09-621-976-16656	Sequence 16656, A
11	43.8	2.0	1176	4	US-09-252-991A-547	Sequence 547, App
12	43.8	2.0	1983	4	US-09-252-991A-531	Sequence 531, App
13	43.4	2.0	1596	4	US-09-252-991A-5938	Sequence 5938, Ap
C 14	43.4	2.0	1674	4	US-09-252-991A-5894	Sequence 5894, Ap
C 15	43	2.0	486	4	US-09-489-039A-2746	Sequence 2746, Ap
16	43	2.0	798	4	US-09-489-039A-2833	Sequence 2833, Ap
17	43	2.0	1575	4	US-09-266-965-84	Sequence 84, Appl
18	43	2.0	18034	4	US-09-266-965-75	Sequence 75, Appl
19	42.8	2.0	1971	4	US-09-252-991A-8432	Sequence 8432, Ap
20	42.8	2.0	2535	4	US-09-252-991A-8466	Sequence 8466, Ap
C 21	42.8	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
22	42.6	2.0	2469	4	US-09-252-991A-5349	Sequence 5349, Ap
C 23	42.6	2.0	5103	4	US-09-252-991A-5192	Sequence 5192, Ap
24	42.6	2.0	6876	4	US-09-252-991A-5283	Sequence 5283, Ap
25	42.2	1.9	1027	4	US-09-674-741-9	Sequence 9, Appli
26	42	1.9	687	4	US-09-252-991A-2644	Sequence 2644, Ap
C 27	42	1.9	804	4	US-09-252-991A-2187	Sequence 2187, Ap

ALIGNMENTS

RESULT 1

US-09-620-312D-1032  
; Sequence 1032, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunqing

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Dmanac, Radoje T.

; TITLE OF INVENTION: No. 6569662el Nucleic Acids and

; FILE REFERENCE: 784CIP2B

; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt\_FL\_genes Version 1.0

; SEQ ID NO 1032

; LENGTH: 1501

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (161)..(1468)

US-09-620-312D-1032

Query Match 61.8%; Score 1345.8; DB 4; Length 1501;

Best Local Similarity 99.9%; Pred.No. 0;

Matches 1347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 593 GGAAGACATCTCGAGCGATGAGGTCAACGTGGCCGTGTGGCCAGCCTCTTTGCCGG 652

Db 1 GGAAGACATCTCGAGCGATGAGGTCAACGTGGCTGTGGTGGCCAGCCTCTTTGCCGG 60

QY 653 ACCTCTGTGGCCCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATTCGCAA 712  
Db |||||  
QY 61 ACCTCTGTGGCCCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATTCGCAA 120  
Db |||||  
QY 713 TTACGGCTACGGTAATGTGGCCCTGATGCCCTCAITGAAATGGACCCCTGAGGCCAGTGA 772  
Db |||||  
QY 121 TTACGGCTACGGTAATGTGGCCCTGATGCCCTCAITGAAATGGACCCCTGAGGCCAGTGA 180  
Db |||||  
QY 773 CCTCTCCGGGGGATTTCTGGCGCTCAGAGATGGCTGTGAGAGCTGGGGTTCAGCAATA 832  
Db |||||  
QY 181 CCTCTCCGGGGGATTTCTGGCGCTCAGAGATGGCTGTGAGAGCTGGGGTTCAGCAATA 240  
Db |||||  
QY 833 TACAGGGGGCCGAGGGCTCAGCGTGGGCCCATCTCAGCAGCAGTGCCTCGATATCTT 892  
Db |||||  
QY 241 TACAGGGGGCCGAGGGCTCAGCGTGGGCCCATCTCAGCAGCAGTGCCTCGATATCTT 300  
Db |||||  
QY 893 CTCGCAATAGAAATGGGCTAACTTCTTTTCCACACCGGGGGGATGGCACCTTTGT 952  
Db |||||  
QY 301 CTCGCAATAGAAATGGGCTAACTTCTTTTCCACACCGGGGGGATGGCACCTTTGT 360  
Db |||||  
QY 953 GGAGCTGGCCGAGTGTGGTGTGGACGACCCACCAGCATGGGGAGGTGTGCCCT 1012  
Db |||||  
QY 361 GGAGCTGGCCGAGTGTGGTGTGGACGACCCACCAGCATGGGGAGGTGTGCCCT 420  
Db |||||  
QY 1013 GGCTGACTTCAACCGTGAATGGCAAGTGGACATCTGTATGGCAACTGGAATGGCCCCA 1072  
Db |||||  
QY 421 GGCTGACTTCAACCGTGAATGGCAAGTGGACATCTGTATGGCAACTGGAATGGCCCCA 480  
Db |||||  
QY 1073 CGCCCTCTATCTCAAAATGAGCCCATGGGAAGTCCGCTTCGGGGACATCGCTCAAC 1132  
Db |||||  
QY 481 CGCCCTCTATCTCAAAATGAGCCCATGGGAAGTCCGCTTCGGGGACATCGCTCAAC 540  
Db |||||  
QY 1133 CAAGTTCTCATGCTCCCTCTGTCCGACGGTCTATCAACCGGACATTTGACAAATGACCA 1192  
Db |||||  
QY 541 CAAGTTCTCATGCTCCCTCTGTCCGACGGTCTATCAACCGGACATTTGACAAATGACCA 600  
Db |||||  
QY 1193 GGAGCTGGAGATCTTCTTCAACAACTGCTTACCGCAGTCTCTACGCAACCGCTCTT 1252  
Db |||||  
QY 601 GGAGCTGGAGATCTTCTTCAACAACTGCTTACCGCAGTCTCTACGCAACCGCTCTT 660  
Db |||||  
QY 1253 CGCGCTCATCGTAGAGACGAGACCCCTCATCGAGGAGCTCAATCCCGGCGACGC 1312  
Db |||||  
QY 1313 CTTGAGCCTGAGGGCGGGGACAGGGGTGTGTGACCGACTTCAGCGAGACGGAT 1372  
Db |||||  
QY 721 CTTGAGCCTGAGGGCGGGGACAGGGGTGTGTGACCGACTTCAGCGAGACGGAT 780  
Db |||||  
QY 1373 GCTGGACCTCATCTTGTCCATGGAGTCCATGGCTCAGCGCTGTCTCGGG 1432  
Db |||||  
QY 781 GCTGGACCTCATCTTGTCCATGGAGTCCATGGCTCAGCGCTGTCTCGGG 840  
Db |||||  
QY 1433 CAATCAGGGCTTCAACAACTGGTGGAGTGGTGCCACGACCCCGTTTGGGGCCTT 1492  
Db |||||  
QY 841 CAATCAGGGCTTCAACAACTGGTGGAGTGGTGCCACGACCCCGTTTGGGGCCTT 900  
Db |||||  
QY 1493 TSCCAGGGGAGTAAAGTGTGTCTACACCAAGAGTGGGGCCCACTGAGATCAT 1552  
Db |||||  
QY 901 TSCCAGGGGAGTAAAGTGTGTCTACACCAAGAGTGGGGCCCACTGAGATCAT 960  
Db |||||  
QY 1553 CGACGGGGGCTCAGGCTACTGTGTGAGATGGAGCCCTGGGACACTTTGGCCTGGGAA 1612  
Db |||||  
QY 961 CGACGGGGGCTCAGGCTACTGTGTGAGATGGAGCCCTGGGACACTTTGGCCTGGGAA 1020  
Db |||||  
QY 1613 GGATGAAGCCAGCAGTGTGAGTGAAGTGGCGCAGATGGCAAGATGGTGAAGCGT 1672  
Db |||||  
QY 1021 GGATGAAGCCAGCAGTGTGAGTGAAGTGGCGCAGATGGCAAGATGGTGAAGCGT 1080  
Db |||||  
QY 1673 GGCACGGGGGAGATCAACTCAGTGTGGAGTCTCTACCCCGGGATGGGACACT 1732  
Db |||||  
QY 1081 GGCACGGGGGAGATCAACTCAGTGTGGAGTCTCTACCCCGGGATGGGACACT 1140  
Db |||||  
QY 1733 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAAGGATTTCTCCCGACAGGAAATGGCCATTG 1792

Db 1141 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAAGATTTCCAGCAGGAAATGGCCATTG 1200  
QY 1793 CATGGACACCAATGAATGATCCAGTTCCCAITTCGTGTGCCCTCGAGACAAAGCCCGTATG 1852  
Db 1201 CATGGACACCAATGAATGATCCAGTTCCCAITTCGTGTGCCCTCGAGACAAAGCCCGTATG 1260  
QY 1853 TGTCAACACCTATGGAAGCTACAGTTCGCGGACCAACAGAGTGCAGTCGGGGCTACGA 1912  
Db 1261 TGTCAACACCTATGGAAGCTACAGTTCGCGGACCAACAGAGTGCAGTCGGGGCTACGA 1320  
QY 1913 GCCCAACGAGGATGGCACAGCCTGCGTGG 1941  
Db 1321 GCCCAACGAGGATGGCACAGCCTGCGTGG 1349

RESULT 2  
US-09-023-655-484  
; Sequence 484, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HEREMITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 484:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 789 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: SYNORAB01  
; CLONE: 192279  
; US-09-023-655-484

Query Match 36.2%; Score 788; DB 4; Length 789;  
Best Local Similarity 99.9%; Pred. No. 3.8e-200;  
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1058 CTGGAATGCCCCCACCCTCTATCTGCAATGACACCCATGGGAAGTCCGCTTCOG 1117  
Db 1 CTGGAATGCCCCCACCCTCTATCTGCAATGACACCCATGGGAAGTCCGCTTCOG 60

```
QY 1118 GGACATCGCTCACCCAGTTCTCCATGCTCCCTCCCTGTCGCGAGGTCATCACGGCGA 1177
Db 61 GGACATGCTCACCACCAAGTTCTCCATGCTCCCTCCCTGTCGCGAGGTCATCACGGCGA 120
QY 1178 CTTTGAACAATGACGAGGCTGGAGATCTTCTTCAACAACATTCGCTACCGCAGCTCTC 1237
Db 121 CTTTGAACAATGACGAGGCTGGAGATCTTCTTCAACAACATTCGCTACCGCAGCTCTC 180
QY 1238 AGCCAAACCGCTCTTCGCGCTCATCCGATGAGAGACGAGACCCCTCATCGAGGAGCT 1297
Db 181 AGCCAAACCGCTCTTCGCGCTCATCCGATGAGAGACGAGACCCCTCATCGAGGAGCT 240
QY 1298 CAATCCCGGACCGCTTGGAGCTGAGCGGCGGGGACACGGGGTGTGTCGAGCTT 1357
Db 241 CAATCCCGGACCGCTTGGAGCTGAGCGGCGGGGACACGGGGTGTGTCGAGCTT 300
QY 1358 CGACGAGACGGGATGCTGGACCTCATCTTCTCCATGGAGTCCATGGCTCAGCCGCT 1417
Db 301 CGACGAGACGGGATGCTGGACCTCATCTTCTCCATGGAGTCCATGGCTCAGCCGCT 360
QY 1418 GTCCGCTTTCGGGGCAATCAGGCTTCAACAACTGGCTGCGAGTGTGCCACGAC 1477
Db 361 GTCCGCTTTCGGGGCAATCAGGCTTCAACAACTGGCTGCGAGTGTGCCACGAC 420
QY 1478 CCGGTTTGGGCGCTTTCGCGGAGCTAAGTGTGCTCTACACCAAGAGTGGGGC 1537
Db 421 CCGGTTTGGGCGCTTTCGCGGAGCTAAGTGTGCTCTACACCAAGAGTGGGGC 480
QY 1538 CCACCTGAGGATCATCGACGGGGCTCAGGCTACTGTGTGAGATGAGCCGTTGCCACA 1597
Db 481 CCACCTGAGGATCATCGACGGGGCTCAGGCTACTGTGTGAGATGAGCCGTTGCCACA 540
QY 1598 CTTTGGCTTGGGGAAGATCAAGCAGCAGTGTGAGGTGAGGTGAGTGGCCAGATG 1657
Db 541 CTTTGGCTTGGGGAAGATCAAGCAGCAGTGTGAGGTGAGGTGAGTGGCCAGATG 600
QY 1658 GGTGAGCGGACGTCGCGCGGGGAGTGAATCAGTGTGAGATGAGTGTGAGTGTGAGT 1717
Db 601 GGTGAGCGGACGTCGCGCGGGGAGTGAATCAGTGTGAGATGAGTGTGAGTGTGAGT 660
QY 1718 GGATGAGGACACATCTCAGGACCCAGCCGCTGAGTGTGGCCAGGATTTCTCCAGCA 1777
Db 661 GGATGAGGACACATCTCAGGACCCAGCCGCTGAGTGTGGCCAGGATTTCTCCAGCA 720
QY 1778 GGAATATGGCATTCGATGACACCAATGATGATCCAGTTCCTGTCGCTGCTG 1837
Db 721 GGAATATGGCATTCGATGACACCAATGATGATCCAGTTCCTGTCGCTGCTGCTG 780
QY 1838 AGACAAGCC 1846
Db 781 AGACAAGCC 789

RESULT 3
US-09-023-655-209
; Sequence 209, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 707 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SYNORAB01
CLONE: 079378
US-09-023-655-209

Query Match 5.0%; Score 109; DB 4; Length 707;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1834 CTCGAGACAAGCCGCTATGTGTCAACACCTATGGAAGCTACAGTGCAGGACCAAGA 1893
Db 1 CTCGAGACAAGCCGCTATGTGTCAACACCTATGGAAGCTACAGTGCAGGACCAAGA 60

QY 1894 AGTGAGTGGGGCTACGAGCCCAACGAGGATGCGACAGCTGCGTGGG 1942
Db 61 AGTGAGTGGGGCTACGAGCCCAACGAGGATGCGACAGCTGCGTGGG 109

RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.7%; Score 58.2; DB 3; Length 4403765;
Best Local Similarity 41.7%; Pred. No. 0.0012;
Matches 363; Conservative 0; Mismatches 508; Indels 0; Gaps 0;

QY 578 CCGCAATAACCGGTGGGAGACATCTCGAGCATGAGTCAACGTCGCGCTGTGTGGC 637
```

```
Db 3733540 CGCGGAACGATTCGACCGCGCACCTTGGTTGCCCGGTGCGGATGGCGGTATGGG 3733481
QY 638 CAGCCTCTTTGCGGACGCTCTGTGGCTGTGGACAGAAAGGCTCTTGGACGCTACTC 697
Db 3733480 CGGCAACGCGCGCAAGGCGGTGACGGGGTGTGGCGCGCGCGGCGGACCGCGTCCG 3733421
QY 698 TATCTACATTCGCAATTAGCCTACGCTAATGTGGGCCCTGATGCCCTCATTTGAATGGA 757
Db 3733420 CGGCAACGCGCGCAAGGCGGAGACGGCGCGCGCGGTGACCGCGCAAGCGCGGTGA 3733361
QY 758 CCTCAGGCGCACTGACCTCTCCGCGGCAATTCGCGCTTCAGAGATGTGGCTGCTGAGCG 817
Db 3733360 TGGCGAGCGGTGCGCGCGGTGATGTGACCTTGGCCGTCAACAGGCTGCGCGGTGA 3733301
QY 818 TGGGGTTCAGCAATATACAGGGGGCGGAGCGCTCAGCGTGGGCCCCCATCTCAGCAGCAG 877
Db 3733300 CGGCGGCAACGCGCGGTGAAGTGGCGTTGCGGCAAGGTGGGCGCGCGGTGTTAGCGC 3733241
QY 878 TGCCTCGGATATCTTCTGGCAATAGAAATGGGCTTAATCTTCTTTTCCAAACCGGGG 937
Db 3733240 GAACCGCGCCCTGAACGCTTCGGCGCGGCGGCAACGCGCACCGCGCCACAGCGGCGCA 3733181
QY 938 CGATGCGACCTTTGTGGACGCTCGCGGCGAGTGTGGTGTGGACGACCCCGCACGACATCG 997
Db 3733180 CGGTGGCAACGCGGCTGCGCGCGCACCCCGCTGCGGGAGAAACGGCGGCGCGG 3733121
QY 998 GCGAGTGTGCGCTGCTGACTTCAACCGTGTGCAAAAGTGGACATCGTCTATGGCA 1057
Db 3733120 TGGTAACGCGCGCATGCGGCGGTGCGGTGCGTAACGCGGTGCGGCTGCGCGGCGGAAA 3733061
QY 1058 CTGGAATGCGCGCCACCGCTCTATCTGCAATGACGCCCATGGAAGTCCGCTTCG 1117
Db 3733060 TGGCGTTCGCGCGCACCGGCTTGGCCCTCAACGCGCGCAACGCGCGCAACGCGGCA 3733001
QY 1118 GGACATTCGCTCAACCAAGTTCCTCATGCTCCCTGTCGCGACGCTCATCACCAGCGA 1177
Db 3733000 CGGCAACGCGGATCGCGCGCGGCAACGCGCGGCGGCGGCAAGCGCGCAACGCGG 3732941
QY 1178 CTTTGACAAATGACGAGCTGAGATCTTTCTTCAACAACTTGCCTACGCGAGCTCTCTC 1237
Db 3732940 CGCGGAGCGCAACGCGCAAGACTTCTCCGCTCGCGCAATGCGCGCAATGCGCGA 3732881
QY 1238 AGCCACCGCTTTCCGCGTCTATCGTAGAGACGAGACCGGACCCCTCATCGAGAGCT 1297
Db 3732880 CGGCAACGCGCGCAACGCGCGGATCGCGCGCAAGGCTGAGCGCTTTCGCCACGTTCCG 3732821
QY 1298 CAATCCCGCGGACGCTTGGAGCTTGAGGCGCGGCGCACAGGGGCTGTGTGACCGACTT 1357
Db 3732820 TAAGCGCGCAACGCGGCTGCGCGGCGCAACGCGCGCAATGTGGCGTTGCGCGCGAG 3732761
QY 1358 CGACGAGACGGAATGCTGGAAGCTCATCTTGTTCCTCCATGAGAGTCCATGCTCAGCGCT 1417
Db 3732760 TGGGCGCGCGCGCAAGGCGGCAATTCAGCCATGAAGGCTGCGACCGCGCGCATGGCAC 3732701
QY 1418 GTCCGTCTTCGCGGCAATCAGGCTTCAAC 1448
Db 3732700 CGCACCCACGCGCGGCTGACGCGCGCAAC 3732670
```

## RESULT 5

```
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
```

```
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Kv
US-09-103-840A-1

Query Match 2.5%; Score 54.8; DB 3; Length 4411529;
Best Local Similarity 41.6%; Pred. No. 0.0093;
Matches 350; Conservative 0; Mismatches 492; Indels 0; Gaps 0;

QY 607 GCGATGAGGTCAACGTGGCCCGTGTGGCCAGCCTCTTTGCCGACGCTCTGTGGCCT 666
Db 3741407 GCGAGCGGCGCTCGCGCGCAAGGCGGTGCGCGCGGTAGCGCGCGCGCGGAAACG 3741348
QY 667 GTGTGACAGAAAGGCTCTGACGCTACTCTATCTACATTTGCAATTACGCTTACGGTA 726
Db 3741347 GATTCGACGCGCCACCTTGGGTTCGCCGCTGCGATGGCGGCAACGCG 3741288
QY 727 ATGTGGCCCTGATGCCCTCATTTGAAATGAGACCTTGAGCGCAGTGAACCTTCCCGG 786
Db 3741287 GCAAGGCGGTGACGCGCGCAAGGCGGTGATGGCGGAGCGCGTGCCTGGTGTGA 3741228
QY 787 TTCTGCGCTCAGAGATGGCTGCTGAGGCTGGGTGAGCAATAATACAGGGGCGCGAG 846
Db 3741227 CCTTGGCGTCAACAGGCTGCGCGGTGACGCGGCAACGCGGTGAGTGGCGTTG 3741168
QY 847 GCGTCAGCGTGGGCGCCCTCCTCAGCAGTGCCTTCGATATCTTCTGCGCAATGAGA 906
Db 3741167 GCGCAAGGCTGGGCGCGGTGTAGCGCAACCGCGCTTGAACGCTTGGCGCGGG 3741108
QY 907 ATGGCGCTAACTTCTCTTTTCCAAACGCGGCGATGCGACCTTTGAGAGCTGCGGCG 966
Db 3741107 CGAACGCGCACCGCGCGCACAGCGCGCAACGCGTGGCAACGAGGTGCGCGCGCAC 3741048
QY 967 GTGTGCTGTGACGACCGCGCGCGAGTGTGCGCGGAGTGTGCGCTTGGCTGACTTCA 1026
Db 3741047 CCACGCTCGCGGAGAAACGCGGCGCGGTGTGTAACGCGGCGCATGCGGCTCGT 3740988
QY 1027 GTGATGGCAAGTGGACATCTGTATGCAATGGAATGGCCCCCACCGCTTCTATTCG 1086
Db 3740987 GTAACGCGGTGCGGCTGTGCGCGGAAATGCGCTGCGCGCACCGGCTTGGCTTCA 3740928
QY 1087 AAATGAGACCCNTGGGAAGTTCGCTTCGGGACATCGCTTCAACCAAGTTCTCCATGC 1146
Db 3740927 ACGGCGCAACGCGCGCAACGCGGCTGCGCGCAACGCGGATCGGCGCGCGCACGG 3740868
QY 1147 CTTCCCTCTGCGCAGCTCATCAGCGCGACTTTTGCATATGACAGGAGCTGGAGATCT 1206
Db 3740867 GCGGCGACGCGGCAAGGCGGCAACGCGGCGCGCGCGCGCAACGCGCAAGACTTCTCCG 3740808
QY 1207 TCTTCAACAACTGCTTACCGCAGCTCTCTCAGCCAAACCGCTTTCGCGTTCATCGTA 1266
Db 3740807 CGTCCGCAATGCGCAATGCGGACAGCGCGCAACGCGCGCAACGCGGCGCATCGCG 3740748
QY 1267 GAGAGCAGGAGACCCCTCATTCAGAGAGCTCAATCCCGCGAGCGCTTGGAGCTCAGG 1326
Db 3740747 GCAAGGCTGTGACGCTTCGCCACGCTTGAAGCGCGCAACGCGGCTGCGCGGCG 3740688
QY 1327 GCGGCGGCAAGGCGGTGTGATCGGCTTCCAGCGAGACGCGGATGTGGAATCTCATCT 1386
Db 3740687 ACGGCGCAATGTGGCGTTGCGGCGAGGCTGGGCGCGCGCAAGGCGCGCATTCAG 3740628
QY 1387 TGTCCCATGAGAGTCAATGGCTCAGCGCTGTCGCTTTCGCGGCAATCAGGCTTCA 1446
Db 3740627 CCATGAAGGCTGCGCGCGCGGATGGACCGCACCGCGCGGTGACGCGCGCA 3740568
QY 1447 AC 1448
Db 3740567 AC 3740566
```

## RESULT 6

US-09-252-991A-10929  
; Sequence 10929, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10929  
; LENGTH: 1221  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10929

Query Match 2.2%; Score 47.8; DB 4; Length 1221;  
Best Local Similarity 50.5%; Pred. No. 0.0078;  
Matches 146; Conservative 0; Mismatches 137; Indels 6; Gaps 1;  
  
QY 1097 CCATGGGAAGTCCGCTTCGGGACATCGCCTCACCCAAAGTTCTCCATGCGCTCCCTCGT 1156  
Db 345 CGACGGCAACATCGCGTCTGTGAAATCAGCAGGACGATCGCTGGTCCGCGACACCG 404  
  
QY 1157 CCGCAGGTCATCACCGCCGACTTTGACATGACAGGAGCTGGAGATCTTCTTCAACAA 1216  
Db 405 CCCGACCTTCGTCATCGACGACAAAGGCGGATGACGGCGGCTGACCTGGGGCTTCAAC-- 462  
  
QY 1217 CATTGCTACCGCAGCTCTCAGCCAAACCGCTTCCGGGTATCCGTAGAGAGCAGG 1276  
Db 463 ----GCTGGGGCGGCTTCGAAAGCGGCTGTACTTCCCTGGCAGCGCAGCAGG 518  
  
QY 1277 AGACCCCTCATCGAGGAGCTCAATCCGCGGACGCTTGGAGCCTGAGGCGCGGGGCAC 1336  
Db 519 GGCACGCAAGATCCTCGAGATCGACGGCGGCGCGCTACCGCACCGACACTTCGTCCT 578  
  
QY 1337 AGGGGTGTGGTACCGACTTCGACGAGACGGGATCTGGACCTCATC 1385  
Db 579 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTGATCACCACC 627

## RESULT 7

US-09-252-991A-11061  
; Sequence 11061, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11061  
; LENGTH: 1269  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11061

Query Match 2.2%; Score 47.8; DB 4; Length 1269;  
Best Local Similarity 50.5%; Pred. No. 0.008;

Matches 146; Conservative 0; Mismatches 137; Indels 6; Gaps 1;  
  
QY 1097 CCATGGGAAGTCCGCTTCGGGACATCGCCTCACCCAAAGTTCTCCATGCGCTCCCTCGT 1156  
Db 316 CGACGGCAACATCGCGTCTGTGAAATCAGCAGGACGATCGCTGGTCCGCGACACCGG 375  
  
QY 1157 CCGCAGGTCATCACCGCCGACTTTGACATGACAGGAGCTGGAGATCTTCTTCAACAA 1216  
Db 376 CCCGACCTTCGTCATCGACGACAAAGGCGGATGACGCGGCTCGACTGGGGCTTCAAC-- 433  
  
QY 1217 CATTGCTACCGCAGCTCTCAGCCAAACCGCTTCCGGGTATCCGTAGAGAGCAGG 1276  
Db 434 ----GCTGGGGCGGCTTCGAAAGCGGCTGTACTTCCCTGGCAGCGCGACGACGAGT 489  
  
QY 1277 AGACCCCTCATCGAGGAGCTCAATCCGCGGACGCTTGGAGCCTGAGGCGCGGGGCAC 1336  
Db 490 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCGCTACCGCACCGACGACTTCGTCCT 549  
  
QY 1337 AGGGGTGTGGTACCGACTTCGACGAGACGGGATCTGGACCTCATC 1385  
Db 550 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTGATCACCACC 598

## RESULT 8

US-09-252-991A-11134/C  
; Sequence 11134, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 11134  
; LENGTH: 1641  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11134

Query Match 2.2%; Score 47.8; DB 4; Length 1641;  
Best Local Similarity 50.5%; Pred. No. 0.0092;  
Matches 146; Conservative 0; Mismatches 137; Indels 6; Gaps 1;  
  
QY 1097 CCATGGGAAGTCCGCTTCGGGACATCGCCTCACCCAAAGTTCTCCATGCGCTCCCTCGT 1156  
Db 1399 CGACGGCAACATCGCGTCTGTGAAATCAGCAGGACGATCGCTGGTCCGCGACACCGG 1340  
  
QY 1157 CCGCAGGTCATCACCGCCGACTTTGACATGACAGGAGCTGGAGATCTTCTTCAACAA 1216  
Db 1339 CCCGACCTTCGTCATCGACGACAAAGGCGGATGACGCGGCTCGACTGGGGCTTCAAC-- 1282  
  
QY 1217 CATTGCTACCGCAGCTCTCAGCCAAACCGCTTCTCCGGGTATCCGTAGAGAGCAGG 1276  
Db 1281 ----GCTGGGGCGGCTTCGAAAGCGGCTGTACTTCCCTGGCAGCGCGACGACGAGT 1226  
  
QY 1277 AGACCCCTCATCGAGGAGCTCAATCCGCGGACGCTTGGAGCCTGAGGCGCGGGGCAC 1336  
Db 1225 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCGCTACCGCACCGACGACTTCGTCCT 1166  
  
QY 1337 AGGGGTGTGGTACCGACTTCGACGAGACGGGATCTGGACCTCATC 1385  
Db 1165 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTGATCACCACC 1117

## RESULT 9

US-09-252-991A-10991  
; Sequence 10991, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10951
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10991

Query Match      2.1%; Score 46.8; DB 4; Length 501;
Best Local Similarity 50.7%; Pred. No. 0.0089;
Matches 142; Conservative 0; Mismatches 132; Indels 6; Gaps 1;

QY 1097 CCATGGGAGGTCGGCTCCGGGACATCGCCTCACCCAAAGTTCTCCATGCCCTCCCTCTGT 1156
DB 227 CGACGGCAACATCCGGTCTGTGGAAATACAGCAGCAGATGCTGGTCCGGACACCGG 286
QY 1157 CGCAGCGTCATCACCGCGACTTTGACATGACACAGGAGCTGGAGATCTTCTCAACAA 1216
DB 287 CCGGACCTTCGTATCGACGACAAAGGGCGATGACGCGGCTGACTGGGGCTTCAAC-- 344
QY 1217 CATTCGCTACCGAGCTCTCAGCCAAACCGCTCTCCCGGTCATCCGTAGAGACACGG 1276
DB 345 ----GCTGGGGCGGCTTCAAGGGCGCTGTACTTCCCTTGCGACGCGACACAGGT 400
QY 1277 AGACCCCTTCATCGAGAGCTCAATCCCGCGACGCTTTGGAGCCTTGAGCGCGGGGCAC 1336
DB 401 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCCCTACCGCACCGACGACTTCGTCT 460
QY 1337 AGGGGTGTGTGACCGACTTCGACGAGACGGAGTGTG 1376
DB 461 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCAAGCTG 500

RESULT 10
US-09-621-976-16656
; Sequence 16656, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16656
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16656

Query Match      2.0%; Score 43.8; DB 4; Length 430;
Best Local Similarity 16.5%; Pred. No. 0.051;
Matches 51; Conservative 123; Mismatches 135; Indels 0; Gaps 0;

QY 915 AACTTCCTTTTCACAAACGGGGGATGGCAGCTTTGTGACGCTCGGGCAGTGTGCTGT 974
DB 59 AMCCYIKKKKSGSRAMCCCTCYKSCSSSTGSGYTTKRAMMWRKRRSCYTSRRMYI 118
QY 975 GTGGACGACCCCAACCAAGCATGGGCGAGGTGTGCGCCTGGCTCACTTCAACCGTGTATGC 1034
```

```
DB 119 YWRSYWMRSMWKSGCCSCSGCYKKKKKKGGSCMSYWMCCYYKRRRMMWKGS 178
QY 1035 AAATGGACATCGTCTATGGAACTGGAAATGGCCCCACCGCTCTATCTGCAATGAGC 1094
DB 179 CMNYTKRMRMRMCCCMRRRRMRMRMCMWKGSYTYCYKSSMMCMARWKRARGKKRMC 238
QY 1095 ACCCATGGGAAGTCCGCTTCGGGACATCGCTCACCCAAAGTTCTCCATGCCCTCCCT 1154
DB 239 CYTKGGGRMMRYCCMRKGRACCTGTTCACTGCCGTGTCTGCCAGAGGCTTCACC 298
QY 1155 GTCCGACACGTCATCACCCCGACTTTGACAATGACCAAGAGCTGGAGATCTTCTTCAAC 1214
DB 299 TACCAGCGCATGCKRAMCSCMMMKRARKKKYMMAMSRMSKMYMARRRRGMMCYKSM 358
QY 1215 AACATTGCC 1223
DB 359 MSKMYKSS 367

RESULT 11
US-09-252-991A-547
; Sequence 547, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 547
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-547

Query Match      2.0%; Score 43.8; DB 4; Length 1176;
Best Local Similarity 47.6%; Pred. No. 0.089;
Matches 166; Conservative 0; Mismatches 177; Indels 6; Gaps 1;

QY 1112 CTTCCGGGACATCGCTCACCCAAAGTTCTCCATGCCCTCCCTGTCCGACGGTATCAC 1171
DB 180 CTTCTGCTTCATGGCGGACGACCGCATCTCTTCTTCAAGCCGGGCGAGTTCGTAC 239
QY 1172 CGCCGACTTTGACAATGACCAAGAGCTGGAGATCTTCTTCAACAACATTGCCCTACCGCAG 1231
DB 240 CTTGAGCTGGAGATCGACGGGAGCGGTGATGCGCTCTTACACCATCTCCAGCTCGCC 299
QY 1232 CTCCTCAGCCAAACCGCTCTTCGGGTCTATCCGTGATAGAGACGAGAGACCCCTCATCGA 1291
DB 300 CTCGGTGCCTACAGCTTCTCCATCACCATCAAGCGGTGCCGGGGGGGGTTCCTCAA 359
QY 1292 GGAGCT-----CAATCCCGGCGAGCGCTTGAGCTGAGGGCGGGGGGACAGGGGGTGT 1345
DB 360 CTGGCTGCACGACAACTCAAGGAAGGCCAGGAGCTACCGGTGACGGTCCGGTCCGCCCT 419
QY 1346 GTGACCGACTTCGACGGGAGACGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCAT 1405
DB 420 GTTCAACGCCATCGACTTCCCGGCGGACAAAGTGTGTCTTCTCTCCGGGGGTTCGGCAT 479
QY 1406 GGCTCAGCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAAAC 1454
DB 480 CACCCCGGTGATGTCATCGCGCTGTGTTCTTCGACACCAACGCCAAC 528

RESULT 12
US-09-252-991A-531
```



; Sequence 531, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 531  
; LENGTH: 1983  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-531

Query Match 2.0%; Score 43.8; DB 4; Length 1983;  
Best Local Similarity 47.6%; Pred. No. 0.12;  
Matches 166; Conservative 0; Mismatches 177; Indels 6; Gaps 1;  
QY 1112 CTTCCGGGACATCGCTCACCAAGTTCTCCATGCTCCCTCCCTGTCGCCAGCGTTCATCAC 1171  
DB 205 CTTCTGCTTCATGCGCGACGAGCGGATCTGTTCTTCAAGCGCGCGGATTCGTCAC 264  
QY 1172 CGCGACTTTGCAATACGAGGAGCTGGAGATCTTTCAACAACATTTGCTACCGGAG 1231  
DB 265 CTTGGAGTGGAGATCGACGCGGAGCGGTGATCGCTCTACCACTTCAGCTCGCC 324  
QY 1232 CTTCTCAGCAACCGCTCTTCCGGGTATCGGTAGAGACGAGACCCCTCATCGA 1291  
DB 325 CTTGGTCCCTACAGCTTCTCCATCACCATCAAGCGGTGCGGCGGGTTTCAA 384  
QY 1292 GGAGCT-----CAATCCGCGAGCGCTTGAGCTGAGCGCTGAGCGCGGCGAGGGTGT 1345  
DB 385 CTGCTGCACGACAACTCAAGGAGGCGAGGACTACCGGTGACCGTCCGGTCCGCT 444  
QY 1346 GGTGACCGACTTCGACGAGACGGGATGCTGGAGCTCATCTTGTCCCATGGAGATCCAT 1405  
DB 445 GTTCAAGCCATGACTTCCGCGCGACAAAGTGTCTGTTCTCTCCGGCGGGTCCGAT 504  
QY 1406 GGCTCAGCGCTGCTCCCTCTTCCGGGCAATCAGGGCTTCAACAAC 1454  
DB 505 CACCCCGGTGATGTCATGCGCGCTGTTCTTCGACACCAACGCCAAC 553

RESULT 13  
US-09-252-991A-5938  
; Sequence 5938, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5938  
; LENGTH: 1596  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5938

Query Match 2.0%; Score 43.4; DB 4; Length 1596;  
Best Local Similarity 45.3%; Pred. No. 0.13;

Matches 158; Conservative 0; Mismatches 191; Indels 0; Gaps 0;  
QY 1083 CTGCAATGAGCACCCATGGGAAGTTCGCTTCCGGGACATCGCTCACCCAGTTCTCC 1142  
DB 1030 CTGACGACAGCGCGCGGAGTGGCCCTGGCTGGTCAATGCGGAACGCAAGCTGTG 1089  
QY 1143 ATGCCCTCCCTGTCCGACCGGTCAATCACCGCGGACTTTGCAATGACGAGAGTGGAG 1202  
DB 1090 GTGAGAGCCCTGGCCAGCGGGGATCGATTTCGCCCTCGGCTACGACGAGAACAGAG 1149  
QY 1203 ATCTTCTTCAACAATTTGCTACCGAGTCTCTCAGCAACCGCTTTCGCGTCACTC 1262  
DB 1150 CGCTTCCCGAAGGATCCAGCCCATGACTGGTTTCGCCGACCGTACGTGGTGTAGG 1209  
QY 1263 CGTAGAGACGAGAGACCCCTCATCGAGAGCTCAATCCCGGAGCGCTTGGAGCT 1322  
DB 1210 CGCGCGACACCCACGCTTCCCGGGCGCGAGCTGGAGGGTACTTGGCCGAAAGG 1269  
QY 1323 GAGGCGCGGACACAGGGGTGTGTGACCGACTTCGACGAGAGCGGATGTCGACCTC 1382  
DB 1270 CATGCGGTGTGACGCCGTGGAACGAGACAGCGGGGTGATCGACCGCTGCTGCCCG 1329  
QY 1383 ATCTTGTCCCATGAGAGTCCATGGCTCAGCGCTGCTCCGTTCCGGG 1431  
DB 1330 TCCGCGCTGCTGCGGAAGTGGCGGTGCACTGCCGAGCGTGTGGCGG 1378

RESULT 14  
US-09-252-991A-5894/c  
; Sequence 5894, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5894  
; LENGTH: 1674  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5894

Query Match 2.0%; Score 43.4; DB 4; Length 1674;  
Best Local Similarity 45.3%; Pred. No. 0.14;  
Matches 158; Conservative 0; Mismatches 191; Indels 0; Gaps 0;  
QY 1083 CTGCAATGAGCACCCATGGGAAGTTCGCTTCCGGGACATCGCTCACCCAGTTCTCC 1142  
DB 750 CTGACGACAGCGCGCGGAGTGGCCCTGGCTCAATGCGGAACGCAAGCTGTG 691  
QY 1143 ATGCCCTCCCTGTCCGACCGGTCAATCACCGCGACTTTGCAATGACGAGAGTGGAG 1202  
DB 690 GTGAGAGCCCTGGCCAGCGGGGATCGATTTCGCCCTCGGCTACGACGAGAACAGAG 631  
QY 1203 ATCTTCTTCAACAATTTGCTACCGAGCTCTCAGCAACCGCTTTCGCCGTCATC 1262  
DB 630 CGCTTCCCGAAGGATCCAGGCCCATGACTGGTTCCCGACCGCTACGTGGTGTAGG 571  
QY 1263 CGTAGAGACGAGACCCCTCTCATCGAGGAGCTCAATCCCGGCGACGCTTGGAGCT 1322  
DB 570 CGCGCGACACCCACGCTTCCCGGGCGCGGAGCTTACCTGGCCGAAAGG 511  
QY 1323 GAGGCGCGGACACAGGGGTGTGTCGCGACTTCGACGAGAGCGGATGCTGGAGCTC 1382  
DB 510 CATGCGGTGTGACCGCGGTGGAACGAGGACAGCGGGGTGATCGACCGGCTGTGGCGCG 451

QY 1383 ATCTTGTCCTCATGAGAGATCCATGCTCAGCGCTGTCCGTCTTCGGG 1431  
Db 450 TCCGGCTCGCTCGGAAAGTGCGGTGCAACTGCCGACGGTCTGCGG 402

RESULT 15

US-09-489-039A-2746/c  
; Sequence 2746, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2746  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2746

Query Match 2.0%; Score 43; DB 4; Length 486;  
Best Local Similarity 45.4%; Pred. No. 0.09;  
Matches 154; Conservative 0; Mismatches 185; Indels 0; Gaps 0;  
QY 265 ACAGTAATCCACCCAGCTCAACTATGTTGGTGGCAGTTACTGATGTGGACCATGATGGG 324  
Db 342 ACAGTCAGCCACCCACCGGTGGGTGTTGATGATGAGGATCGGCCACCGGTGGC 283  
QY 325 ACTTTGAGATCGTGTGGGGGTACAATGGACCCCAACCTGTTCTGAATATGACCGGG 384  
Db 282 GCACACGACACCTGCCGAGTACAAAGCCGCGCGCGATGCCGAGACGCTGTGG 223  
QY 385 CCAGAGCGGTGGTGAACATCGCGTCTGATGAGCGCAGTCACTTACTAGCGCTGC 444  
Db 222 CCAGATGCCGCGCTGCGGGCGCTTCGACAGCGGGGATCCGCTGTGGCCTCGC 163  
QY 445 GGCACCGCAGGGGACGCCATTGGGTACAGCTCGACATCGACGGGACGGCGGG 504  
Db 162 CGGCAGCAGAGCAGACGATCTGCCGCCACCTGGCGGTGAGGTGGCCAGCGGGCC 103  
QY 505 AGGAGATCTACTTCTCAACACCAATATGCTTCTCGGGGTGGCCACGTAACCCACA 564  
Db 102 ATCAGGCCACCATGCTCTCCACCGACAAAGGCTACTGCCAGTGTCTGTGCCCATCC 43  
QY 565 AGTTGTTCAAGTTCGGCANTAACCGTGGGAAGCATCC 603  
Db 42 GCATCCCGGACTACTTTTCAAGCGCTGGTGGACGCC 4

Search completed: August 9, 2004, 21:19:09  
Job time : 198 secs



QY	371	GAAATATGACCGGGCCCAAGAGCGGTGGTGAACATGCGGTGCGATGAGCGAGCTCACC	430	QY	1451	CAACTGGCTGCGAGTGGTGCCACGACCCGGTGTGGGGCCCTTTGGCGGGAGCTAAGGT	1510
Db	283	GAAATATGACCGGGCCCAAGAGCGGTGGTGAACATGCGGTGCGATGAGCGAGCTCACC	342	Db	1363	CAACTGGCTGCGAGTGGTGCCACGACCCGGTGTGGGGCCCTTTGGCGGGAGCTAAGGT	1422
QY	431	CTACTACGGCTGCGGACCGGACGAGGGAACGCCATTGGGGTCACAGCCTCGACATCGA	490	QY	1511	CGTGTCTACACCAAGAAGAGTGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTA	1570
Db	343	CTACTACGGCTGCGGACCGGACGAGGGAACGCCATTGGGGTCACAGCCTCGACATCGA	402	Db	1423	CGTGTCTACACCAAGAAGAGTGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTA	1482
QY	491	CGGGACCGCCGGGAGGAGATCTACTTCTCAACCAACCAATAATATGCTTCTCGGGGGTGGC	550	QY	1571	CTGTGTGAGATGAGACCCGTGGACACTTTGGCTGGGGAAGGATGAAGCCAGCAGTGT	1630
Db	403	CGGGACCGCCGGGAGGAGATCTACTTCTCAACCAACCAATAATATGCTTCTCGGGGGTGGC	462	Db	1483	CTGTGTGAGATGAGACCCGTGGACACTTTGGCTGGGGAAGGATGAAGCCAGCAGTGT	1542
QY	551	CACGTACACCGACAAGTTCTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTGAGCGA	610	QY	1631	CGAGGTGACGTCGCCAGATGGCAAGATGCTGAGCCGGAAACGTGGCCAGCGGGGAGATGAA	1690
Db	463	CACGTACACCGACAAGTTCTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTGAGCGA	522	Db	1543	GGAGGTGACGTCGCCAGATGGCAAGATGCTGAGCCGGAAACGTGGCCAGCGGGAGATGAA	1602
QY	611	TGAGGTCAACGTCGGCCGCGGTGGTGGCCAGCCTCTTTGCGGACGCTCTGTGGCCCTGT	670	QY	1691	CTCAGTGTGGAGATCTCTACCCCGGGATGAGGACACATTCAGGACCCAGCCCACT	1750
Db	523	TGAGGTCAACGTCGGCCGCGGTGGTGGCCAGCCTCTTTGCGGACGCTCTGTGGCCCTGT	582	Db	1603	CTCAGTGTGGAGATCTCTACCCCGGGATGAGGACACATTCAGGACCCAGCCCACT	1662
QY	671	GGACAGAAAGGGCTCGGACGCTACTCTATCTACATTCGCAATACGCTACGGTAATGT	730	QY	1751	GGAGTGTGSCCAAGGATCTCCACGACGAAATGSCCATTTGCATGGACACCAATGAATG	1810
Db	583	GGACAGAAAGGGCTCGGACGCTACTCTATCTACATTCGCAATACGCTACGGTAATGT	642	Db	1663	GGAGTGTGSCCAAGGATCTCCACGACGAAATGSCCATTTGCATGGACACCAATGAATG	1722
QY	731	GGGCCCTGATGCCCCTCATTTGAAATGGAACCTGAGGCCAGTGACCTCTCCGGGGCAATCT	790	QY	1811	CATCAGTTCCTCANTTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACCTATGGAAG	1870
Db	643	GGGCCCTGATGCCCCTCATTTGAAATGGAACCTGAGGCCAGTGACCTCTCCGGGGCAATCT	702	Db	1723	CATCAGTTCCTCANTTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACCTATGGAAG	1782
QY	791	GGGGCTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAAAATATACAGGGGGCGAGGCGT	850	QY	1871	CTACAGGTCCGGACCAACAAGAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC	1930
Db	703	GGGGCTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAAAATATACAGGGGGCGAGGCGT	762	Db	1783	CTACAGGTCCGGACCAACAAGAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC	1842
QY	851	CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAAATGAGAATGG	910	QY	1931	AGCCTCGTGGGTGGTGGAGCCCTGTGTGAAGTAGTGACACCAACAGTTTGGGAAGAG	1990
Db	763	CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAAATGAGAATGG	822	Db	1843	AGCCTCGTGGGTGGTGGAGCCCTGTGTGAAGTAGTGACACCAACAGTTTGGGAAGAG	1902
QY	911	GCTTAATCTTCTTTTCCAAACCGGGGCGATGSCACTTTGTGGAGCTGCGGCCAGTGC	970	QY	1991	CCTTGTGCTCCTGAATCACTGAATCACTGCTTGAATCAGCGCTGGAATACCTGTTGATC	2050
Db	823	GCTTAATCTTCTTTTCCAAACCGGGGCGATGSCACTTTGTGGAGCTGCGGCCAGTGC	882	Db	1903	CCTTGTGCTCCTGAATCACTGAATCACTGCTTGAATCAGCGCTGGAATACCTGTTGATC	1962
QY	971	TGGTGTGGACGACCCCAACAGCATGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA	1030	QY	2051	AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAACTTCTATTGTATTAGCTATT	2110
Db	883	TGGTGTGGACGACCCCAACAGCATGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA	942	Db	1963	AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAACTTCTATTGTATTAGCTATT	2022
QY	1031	TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCAACCGCTCTATCTGCAAA	1090	QY	2111	AATACATTAAATTTGGGGGTGCTACTTACATAATAAATCCCATTTCCCTTTGAAAAA	2170
Db	943	TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCAACCGCTCTATCTGCAAA	1002	Db	2023	AATACATTAAATTTGGGGGTGCTACTTACATAATAAATCCCATTTCCCTTTGAAAAA	2082
QY	1091	GAGCACCATGGGAAGTGGCTTCCGGGACATCGCTGACCCAGTTCTCCATGSCCTC	1150	QY	2171	AAAAAAA 2177	
Db	1003	GAGCACCATGGGAAGTGGCTTCCGGGACATCGCTGACCCAGTTCTCCATGSCCTC	1062	Db	2083	AAAAAAA 2089	
QY	1151	CCCTGTCCGACCGTCTATCAGCGGACTTTGCAATGACAGAGCTGGAGATCTTCTT	1210				
Db	1063	CCCTGTCCGACCGTCTATCAGCGGACTTTGCAATGACAGAGCTGGAGATCTTCTT	1122				
QY	1211	CACCAACATTGGCTACCGGAGCTCTCTAGCCAAACCGCTCTTCCGGTCAATCGTAGAGA	1270				
Db	1123	CACCAACATTGGCTACCGGAGCTCTCTAGCCAAACCGCTCTTCCGGTCAATCGTAGAGA	1182				
QY	1271	GCAGGAGACCCCTCATTCAGAGCTCAATCCGGGAGCCCTTGAGCCTGAGGCGC	1330				
Db	1183	GCAGGAGACCCCTCATTCAGAGCTCAATCCGGGAGCCCTTGAGCCTGAGGCGC	1242				
QY	1331	GGGACAGGGGTGTGTGACCGACTTTCGACGGAGAGGATGTGACCTCATCTTGTGTC	1390				
Db	1243	GGGACAGGGGTGTGTGACCGACTTTCGACGGAGAGGATGTGACCTCATCTTGTGTC	1302				
QY	1391	CCATGGAGAGTCCATGGCTCAGCGCTGTGCGTCTTCCGGGGCAATCAGGGCTTCAACAA	1450				
Db	1303	CCATGGAGAGTCCATGGCTCAGCGCTGTGCGTCTTCCGGGGCAATCAGGGCTTCAACAA	1362				

RESULT 2  
US-09-765-231A-30  
; Sequence 30, Application US/09765231A  
; Patent No. US20020119452A1  
; GENERAL INFORMATION:  
; APPLICANT: Searle/Monsanto  
; APPLICANT: Phippard, Deborah  
; APPLICANT: Vasanthakumari, Geetha  
; APPLICANT: Dotson, Stanton  
; APPLICANT: Ma, Xiao-Jun  
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,  
; FILE REFERENCE: SO-3221 PR  
; CURRENT APPLICATION NUMBER: US/09/765,231A  
; CURRENT FILING DATE: 2001-01-18  
; NUMBER OF SEQ ID NOS: 82  
; SEQ ID NO 30  
; LENGTH: 2263  
; TYPE: DNA  
; ORGANISM: Homo sapiens















Db 1141 ATCAGAGAGCTCAATCCCGGCGAGCGCCTTGAGAGCCTGAGGGCCGGGGCAGAGGGGGTGTG 1200

Qy 1347 GTGACCGACTTTCGACGGAGAGCGGATGCTGGACCTCATCTTGTGCCCATGGAGATCCATG 1406

Db 1201 GTGACCGACTTTCGACGGAGAGCGGATGCTGGACCTCATCTTGTGCCCATGGAGATCCATG 1260

Qy 1407 GCTCAGCGCGTGTCCGTCTTCGCGGGGCAATCAGGGCTTCAACAACAATCGCTGGAGTG 1466

Db 1261 GCTCAGCGCGTGTCCGTCTTCGCGGGGCAATCAGGGCTTCAACAACAATCGCTGGAGTG 1320

Qy 1467 GTGCCACGACCCGGTGTGGGGCCCTTTCGACGGGAGCTAAGTCTGTCTACACCAAG 1526

Db 1321 GTGCCACGACCCGGTGTGGGGCCCTTTCGACGGGAGCTAAGTCTGTCTACACCAAG 1380

Qy 1527 AAGAGTGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586

Db 1381 AAGAGTGGGGCCCACTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy 1587 CCGGTGGGACACTTGGCTGGGGAAGGATGAAGCAGCAGTGTGGAGTGAAGTGGCCA 1646

Db 1441 CCGGTGGGACACTTGGCTGGGGAAGGATGAAGCAGCAGTGTGGAGTGAAGTGGCCA 1500

Qy 1647 GATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTCTGGAGATC 1706

Db 1501 GATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTCTGGAGATC 1560

Qy 1707 CTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGA 1766

Db 1561 CTCTACCCCGGGATGAGGACACACTTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGA 1620

Qy 1767 TTCTCCAGCAGGAAATGGCCATGATGATGACACCAATGAATGATCCAGTTCGCATTC 1826

Db 1621 TTCTCCAGCAGGAAATGGCCATGATGATGACACCAATGAATGATCCAGTTCGCATTC 1680

Qy 1827 GTGTGCCCTCGAGACAAGCCCGTATGTCTAACACCTATGAAGTACAGTGTGGGACC 1886

Db 1681 GTGTGCCCTCGAGACAAGCCCGTATGTCTAACACCTATGAAGTACAGTGTGGGACC 1740

Qy 1887 AACAGAAGTCAAGTTCGGGGCTACGAGCCCAACGAGGATGGCAGACCTGCGTGGG 1942

Db 1741 AACAGAAGTCAAGTTCGGGGCTACGAGCCCAACGAGGATGGCAGACCTGCGTGGG 1796

RESULT 8

US-09-822-830A-88

; Sequence 88, Application US/09822830A

; Patent No. US20020142952A1

; GENERAL INFORMATION:

; APPLICANT: Genetics Institute, Inc.

; APPLICANT: Wong, Gordon G.

; APPLICANT: Clark, Hilary

; APPLICANT: Fechtel, Kim

; APPLICANT: Agostino, Michael J.

; APPLICANT: Howes, Steven H.

; APPLICANT: Resnick, Richard J.

; APPLICANT: Gulukota, Kamalakar

; APPLICANT: Graham, James R.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS

; FILE REFERENCE: GIN 6402

; CURRENT APPLICATION NUMBER: US/09/822,830A

; CURRENT FILING DATE: 2001-03-29

; PRIOR APPLICATION NUMBER: 60/195,604

; PRIOR FILING DATE: 2000-04-06

; NUMBER OF SEQ ID NOS: 631

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 88

; LENGTH: 1849

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 1810,1813,1814

; OTHER INFORMATION: n=a,c,g, or t

US-09-822-830A-88

Query Match 77.5%; Score 1687.8; DB 9; Length 1849;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1700; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 239 CAACTCAGTTCCTGCTCCTGACTATGACAGTAATCCACCAGCTCAACTATGCTGGC 298

Db 1 CAACTCAGTTCCTGCTCCTGACTATGACAGTAATCCACCAGCTCAACTATGCTGGC 60

Qy 299 AGTTACTGATGTGACCATGATGGGACTTTGAGATCGTCTGCGGGGTACAATGGAGCC 358

Db 61 AGTTACTGATGTGACCATGATGGGACTTTGAGATCGTCTGCGGGGTACAATGGAGCC 120

Qy 359 CAACCTGGTTCCTGAAGTATGACCGGGGCCAGAAAGCGGTGTGTAACATCGCGGTGATGA 418

Db 121 CAACCTGGTTCCTGAAGTATGACCGGGGCCAGAAAGCGGTGTGTAACATCGCGGTGATGA 180

Qy 419 GCGCAGCTCACCTTACTACGGGTGCGGGACCGCAGGGNAACGCCATTGGGGTCAAGC 478

Db 181 GCGCAGCTCACCTTACTACGGGTGCGGGACCGCAGGGNAACGCCATTGGGGTCAAGC 240

Qy 479 CTGCGACATCGACGGGACGCGCGGAGGAGATCTACTTCTCAACACCAATTAATGCGCTT 538

Db 241 CTGCGACATCGACGGGACGCGCGGAGGAGATCTACTTCTCAACACCAATTAATGCGCTT 300

Qy 539 CTCGGGGTGGCCACGTACACCGCAAGTGTTCCTCAAGTTCGCGCAATAACCGGGTGGAAAG 598

Db 301 CTCGGGGTGGCCACGTACACCGCAAGTGTTCCTCAAGTTCGCGCAATAACCGGGTGGAAAG 360

Qy 599 CATCTGAGCGATGAGTCAAGTGGCGCGGTGGTGGCGAGCCTTTGGCGGACGCTC 658

Db 361 CATCTGAGCGATGAGTCAAGTGGCGCGGTGGTGGCGAGCCTTTGGCGGACGCTC 420

Qy 659 TGTGCGCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATTGCGCAATACGC 718

Db 421 TGTGCGCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATTGCGCAATACGC 480

Qy 719 CTACGGTAATGGGCGCTGATGCCCTCATTTGAAATGGAACCTGAGGCGAGTCACTCTC 778

Db 481 CTACGGTAATGGGCGCTGATGCCCTCATTTGAAATGGAACCTGAGGCGAGTCACTCTC 540

Qy 779 CCGGGCATTTCTGGCGCTCAGAGATGGCTGCTGAGGCTGGGGTGCAGCAATATACAGG 838

Db 541 CCGGGCATTTCTGGCGCTCAGAGATGGCTGCTGAGGCTGGGGTGCAGCAATATACAGG 600

Qy 839 GGGCGGAGGCTCAGCGTGGGCCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 898

Db 601 GGGCGGAGGCTCAGCGTGGGCCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 660

Qy 899 CAATGAGAATGGGCTAACTTCTTTTCCAAACCGGGGCGATGGCACCTTTGTGGAGCG 958

Db 661 CAATGAGAATGGGCTAACTTCTTTTCCAAACCGGGGCGATGGCACCTTTGTGGAGCG 720

Qy 959 TGGGCGAGTCTGCTGTGGAGAGCCCAACAGCATGGGCGAGTGTGGCGCTGGCTGGA 1018

Db 721 TGGGCGAGTCTGCTGTGGAGAGCCCAACAGCATGGGCGAGTGTGGCGCTGGCTGGA 780

Qy 1019 CTTCAACCGTGTGCAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCGCCCGCT 1078

Db 781 CTTCAACCGTGTGCAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCGCCCGCT 840

Qy 1079 CTATCTGCAATGAGACCCCATGGGAAGTTCGCTTCGGGACATCGCTCAACCAAGTT 1138

Db 841 CTATCTGCAATGAGACCCCATGGGAAGTTCGCTTCGGGACATCGCTCAACCAAGTT 900

Qy 1139 CTCATGCGCTCCCTGTCCCGCAGTTCATCCGCGAGCTTTCATGACATGACGAGGACT 1198

Db 901 CTCATGCGCTCCCTGTCCCGCAGTTCATCCGCGAGCTTTCATGACATGACGAGGACT 960

Qy 1199 GGAGATCTTCTTCAACCAACTTGTCTACCGAGCTCTCTCAGCCCAACCGCTCTTCCGCGT 1258

Db 961 GGAGATCTTCTTCAACCAACTTGTCTACCGAGCTCTCTCAGCCCAACCGCTCTTCCGCGT 1020

```
QY 1259 CATCCGTAGAGACACGAGACCCCTCATCGAGGAGCTCAATCCCGGACGCGCTTGGG 1318
Db 1021 CATCCGTAGAGACACGAGACCCCTCATCGAGGAGCTCAATCCCGGACGCGCTTGGG 1080
QY 1319 GCCTGAGGGCCGGGACACAGGGGTGTGTGACCCGACTTCCGACGAGACGGGATGCTTGGG 1378
Db 1081 GCCTGAGGGCCGGGACACAGGGGTGTGTGACCCGACTTCCGACGAGACGGGATGCTTGGG 1140
QY 1379 CCTCATCTTGTCCATGAGAGTTCATGGCTCAGCGCTGTGCTTCCCGGGCAATCA 1438
Db 1141 CCTCATCTTGTCCATGAGAGTTCATGGCTCAGCGCTGTGCTTCCCGGGCAATCA 1200
QY 1439 GGGCTTCAACAACTGGCTGGAGTGGTCCACGACCCGGTTTGGGGCCCTTGGCCAG 1498
Db 1201 GGGCTTCAACAACTGGCTGGAGTGGTCCACGACCCGGTTTGGGGCCCTTGGCCAG 1260
QY 1499 GGGAGCTAAGGTCTGTCTTACACCAAGAGAGTGGGGCCCACTGAGGATCATCGACGG 1558
Db 1261 GGGAGCTAAGGTCTGTCTTACACCAAGAGAGTGGGGCCCACTGAGGATCATCGACGG 1320
QY 1559 GGGCTCAGGCTACTGTGTAGATGAGCCCGTGGCACATTTGGCTGGGGAGGATGA 1618
Db 1321 GGGCTCAGGCTACTGTGTAGATGAGCCCGTGGCACATTTGGCTGGGGAGGATGA 1380
QY 1619 AGCCAGCAGTGTGGAGTGCAGTGGCCAGATGCAAGATGGTGAGCCGGAACGTGGCCAG 1678
Db 1381 AGCCAGCAGTGTGGAGTGCAGTGGCCAGATGCAAGATGGTGAGCCGGAACGTGGCCAG 1440
QY 1679 CGGGGAGATGAATCAGTGTGAGATTCCTTACCCCGGGATGAGACACATTCAGGA 1738
Db 1441 CGGGGAGATGAATCAGTGTGAGATTCCTTACCCCGGGATGAGACACATTCAGGA 1500
QY 1739 CCCAGCCCACTGGAGTGTGGCCAAAGATTTCTCCAGCAGAAATGGCCATTCGATGGA 1798
Db 1501 CCCAGCCCACTGGAGTGTGGCCAAAGATTTCTCCAGCAGAAATGGCCATTCGATGGA 1560
QY 1799 CACCAATGAATGCATCCAGTTCCTTCCATTCGGTGTGCCCTCGAGACAAAGCCGCTATGTGCAA 1858
Db 1561 CACCAATGAATGCATCCAGTTCCTTCCATTCGGTGTGCCCTCGAGACAAAGCCGCTATGTGCAA 1620
QY 1859 CACTATGGAAGTACAGGTGCGGACCAACAAAGAGTGCAGTGGGGCTTACGAGCCCAA 1918
Db 1621 CACTATGGAAGTACAGGTGCGGACCAACAAAGAGTGCAGTGGGGCTTACGAGCCCAA 1679
QY 1919 CGAGGATGCCACAGCCTGCGTGG 1941
Db 1680 CGAGGATGCCACAGCCTGCGTGG 1702
```

## RESULT 9

```
US-10-295-027-180
; Sequence 180, Application US/10295027
; Publication No. US2003023350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynné, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-01250005
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
```

```
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 180
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-180

Query Match 54.7%; Score 1408.8; DB 16; Length 2178;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 435 TACGCGCTGCGGACCGGACCGGAGGAAACCGCATTTGGGGTTCACAGCCTCGGACATCGACGGG 494
Db 1 TACGCGCTGCGGACCGGACCGGAGGAAACCGCATCGGGGTTCACAGCCTCGGACATCGACGGG 60
QY 495 GACGCGCGGAGAGATCTATCTTCTCAACACAATAATGCTTCTCGGGGTGGCCACG 554
Db 61 GACGCGCGGAGAGATCTATCTTCTCAACACAATAATGCTTCTCGGGGTGGCCACG 120
QY 555 TACACCGACAAGTTGTTCAAGTTCGGAATAACCGGTGGGAAGACATCTCTGAGCGATGAG 614
Db 121 TACACCGACAAGTTGTTCAAGTTCGGAATAACCGGTGGGAAGACATCTCTGAGCGATGAG 180
QY 615 GTCAACGTGCCCGTGGTGTGGCCAGCCTCTTTGCCGAGCGCTGTGTGGCTGTGTGGAC 674
Db 181 GTCAACGTGCCCGTGGTGTGGCCAGCCTCTTTGCCGAGCGCTGTGTGGCTGTGTGGAC 240
QY 675 AGAAGGGCTCTGACGCTACTCTATCTATCAATTTGCCAATTTACGCTTACGTAATGTGGGC 734
Db 241 AGAAGGGCTCTGACGCTACTCTATCTATCAATTTGCCAATTTACGCTTACGTAATGTGGGC 300
QY 735 CCTGATGCCCTCATTGAATGACCGTGGGCGAGTGCATCTCTCCGGGGCATTTCTGGGC 794
Db 301 CCTGATGCCCTCATTGAATGACCGTGGGCGAGTGCATCTCTCCGGGGCATTTCTGGGC 360
QY 795 CTGAGAGATGTGCTGCTGCTGAGGCTGGGGTTCAGCAAAATATACAGGGGCGGAGCGTCAAG 854
Db 361 CTGAGAGATGTGCTGCTGCTGAGGCTGGGGTTCAGCAAAATATACAGGGGCGGAGCGTCAAG 420
QY 855 GTGGGGCCCATCTCTGACGAGCAGTGCCTCGGATATCTTTCTGCACAATGAGATGGGCCT 914
Db 421 GTGGGGCCCATCTCTGACGAGCAGTGCCTCGGATATCTTTCTGCACAATGAGATGGGCCT 480
QY 915 AACTTCCTTTTCCACAACCGGGGCGATGCACCTTTGTGGAGCTGCGGCGAGTGTGCT 974
Db 481 AACTTCCTTTTCCACAACCGGGGCGATGCACCTTTGTGGAGCTGCGGCGAGTGTGCT 540
QY 975 GTGGAGCAGCCCCCACCAGCATGGGCGAGTGTGCGCCCTGCTGACTTTCAACCGTGTATGGC 1034
Db 541 GTGGAGCAGCCCCCACCAGCATGGGCGAGTGTGCGCCCTGCTGACTTTCAACCGTGTATGGC 600
QY 1035 AAGTGGACATCTCTATGCAACTGGAATGGCCCCCAGCGCTCTATCTGCAAAATGAGC 1094
Db 601 AAGTGGACATCTCTATGCAACTGGAATGGCCCCCAGCGCTCTATCTGCAAAATGAGC 660
```

```
QY 1095 ACCCATGGGAAGTCCGGTTCGGGACATCGCCTCACCCAAAGTTCTCCATGCCCTCCCT 1154
Db 661 ACCCATGGGAAGTCCGGTTCGGGACATCGCCTCACCCAAAGTTCTCCATGCCCTCCCT 720
QY 1155 GTCCGCAAGTTCACACCGCGACTTTGACAAATGACCAAGGAGCTGGAGATCTTCTCAAC 1214
Db 721 GTCCGCAAGTTCACACCGCGACTTTGACAAATGACCAAGGAGCTGGAGATCTTCTCAAC 780
QY 1215 AACATTGCTACCGCAGCTCTCAGCCAAACCGCTCTTCCGCGTCATCCGTCAGAGAGCAC 1274
Db 781 AACATTGCTACCGCAGCTCTCAGCCAAACCGCTCTTCCGCGTCATCCGTCAGAGAGCAC 840
QY 1275 GGAGACCCCTCATCGAGAGCTCAATCCCGGCGAGCGCTTGGAGCCCTGAGGGCGGGG 1334
Db 841 GGAGACCCCTCATCGAGAGCTCAATCCCGGCGAGCGCTTGGAGCCCTGAGGGCGGGG 900
QY 1335 ACAGGGGTGTGGTACCGACTTCGACGAGACGGATGCTGGACCTCATCTTGTCCCAT 1394
Db 901 ACAGGGGTGTGGTACCGACTTCGACGAGACGGATGCTGGACCTCATCTTGTCCCAT 960
QY 1395 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGCAATCAGGGCTTCAACAAAC 1454
Db 961 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGCAATCAGGGCTTCAACAAAC 1020
QY 1455 TGGCTGCGAGTGTGCGACGACCGCGTTTGGGGCTTTGCGAGGGAGCTAAAGTGTGT 1514
Db 1021 TGGCTGCGAGTGTGCGACGACCGCGTTTGGGGCTTTGCGAGGGAGCTAAAGTGTGT 1080
QY 1515 CTCTACACCAAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTG 1574
Db 1081 CTCTACACCAAGAGTGGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTG 1140
QY 1575 TGTGAGATGGAGCCCGTGGCCACATTTGGCCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1634
Db 1141 TGTGAGATGGAGCCCGTGGCCACATTTGGCCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1200
QY 1635 GTGAGCTGGCCAGATGGCAAGATGTGAGCCGGAACTGAGCGAGCGGAGATGAATCTCA 1694
Db 1201 GTGAGCTGGCCAGATGGCAAGATGTGAGCGGGAACTGAGCGAGCGGAGATGAATCTCA 1260
QY 1695 GTGCTGGAGATCCCTACCCCGGGATGAGGACACATTTAGGACCCAGCCCACTGGAG 1754
Db 1261 GTGCTGGAGATCCCTACCCCGGGATGAGGACACATTTAGGACCCAGCCCACTGGAG- 1319
QY 1755 TGTGGCCAAAGGATTTCTCCAGCAGGAAATGGCCATTGCAATGGACCAACAAATGCAATC 1814
Db 1320 -----GACACCAATGAATGCATC 1337
QY 1815 CAGTTCGATTCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACCTATGGAGCTAC 1874
Db 1338 CAGTTCGATTCGTGTGCCCTCGAGACAAAGCCCGTATGTGTCAACACCTATGGAGCTAC 1397
QY 1875 AGGTGCGGACCAACAAGAGTGAGTGGGGCTACGAGCCCAACGAGGATGGCACGCC 1934
Db 1398 AGGTGCGGACCAACAAGAGTGAGTGGGGCTACGAGCCCAACGAGGATGGCACGCC 1457
QY 1935 TGCCTGGG 1942
Db 1458 TGCCTGGG 1465
```

## RESULT 10

```
US-10-188-832-55
; Sequence 55, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
```

```
; FILE REFERENCE: 018501-002330US
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-188-832-55
```

## Query Match

```
Best Local Similarity 64.7%; Score 1408.8; DB 17; Length 2178;
Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;
QY 435 TACGCGCTGCGGACCGGAGGAAAGCCCATTTGGGTCACAGCCTGCGCATCGACGGG 494
Db 1 TACGCGCTGCGGACCGGAGGAAAGCCCATTTGGGTCACAGCCTGCGCATCGACGGG 60
QY 495 GACGGCGGGAGGAGATCTACTTCTCAACACCAATATGCTTCTCGGGGTGGCCACG 554
Db 61 GACGGCGGGAGGAGATCTACTTCTCAACACCAATATGCTTCTCGGGGTGGCCACG 120
QY 555 TACACCGACAAGTTGTTCAAGTTCGCAATAAACCGTGGGAAGACATCCTCAGCCATGAG 614
Db 121 TACACCGACAAGTTGTTCAAGTTCGCAATAAACCGTGGGAAGACATCCTCAGCCATGAG 180
QY 615 GTCACGCTGGCCCTGTGTGTGGCCAGGCTCTTTGCCGACGCTCTGTGCCCTGTGTGGAC 674
Db 181 GTCACGCTGGCCCTGTGTGTGGCCAGGCTCTTTGCCGACGCTCTGTGCCCTGTGTGGAC 240
QY 675 AGAAGGGCTCTGACGCTACTCTATCTACATTCGCAATAGCCTACGTAATGTGGG 734
Db 241 AGAAGGGCTCTGACGCTACTCTATCTACATTCGCAATAGCCTACGTAATGTGGG 300
QY 735 CCTGATGCCCTCATTTGAAATGGACCCCTGAGGCCAGTGACCTCTCCGGGGCAATCTGGG 794
Db 301 CCTGATGCCCTCATTTGAAATGGACCCCTGAGGCCAGTGACCTCTCCGGGGCAATCTGGG 360
QY 795 CTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAAAATATACAGGGGGCCGAGGGCTCAGC 854
Db 361 CTCAGAGATGTGGCTGTGAGGCTGGGGTCAGCAAAATATACAGGGGGCCGAGGGCTCAGC 420
QY 855 GTGGGCCCCATCTTCACGACGAGTGCCTCGGATATCTTCTCGGACAAATGAGAAATGGGCT 914
Db 421 GTGGGCCCCATCTTCACGACGAGTGCCTCGGATATCTTCTCGGACAAATGAGAAATGGGCT 480
QY 915 AACTTCTTTTCCACAAACCGGGGCGATGGCACTTTGTGGACGCTGTGGGCCAGTGTGGT 974
Db 481 AACTTCTTTTCCACAAACCGGGGCGATGGCACTTTGTGGACGCTGTGGGCCAGTGTGGT 540
QY 975 GTGGACGACCCCCACCCAGCATGGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTGATGGC 1034
Db 541 GTGGACGACCCCCACCCAGCATGGGCGAGGTGTGCGCCCTGGCTGACTTCAACCGTGATGGC 600
QY 1035 AAAGTGACATCGTCTATGGCAACTGGAATGGCCCCCAGCCCTCTATCTGCAATGAGC 1094
Db 601 AAAGTGACATCGTCTATGGCAACTGGAATGGCCCCCAGCCCTCTATCTGCAATGAGC 660
QY 1095 ACCCATGGGAAGTCCGCTTCCGGGACATGCGCTCAGCCCAAGTTCCTCATGCCCTCCCT 1154
Db 661 ACCCATGGGAAGTCCGCTTCCGGGACATGCGCTCAGCCCAAGTTCCTCATGCCCTCCCT 720
```



```
QY 1155 GTCCGACAGGTATCATACCGCGCACTTTGACAATGACAGGAGCTGGAGATCTTTCTCAAC 1214
      |||
Db 721 GTCCGACAGGTATCATACCGCGCACTTTGACAATGACAGGAGCTGGAGATCTTTCTCAAC 780
      |||
QY 1215 AACATTGCTTACCGAGCTCTCAGCAACCGGCTTTTCGGCTCATTCGTAGAGCAC 1274
      |||
Db 781 AACATTGCTTACCGAGCTCTCAGCAACCGGCTTTTCGGCTCATTCGTAGAGCAC 840
      |||
QY 1275 GAGACCCCTCATCAGGAGCTCAATCCCGGCGAGCGCTTGGAGCTGAGGCGCGGGC 1334
      |||
Db 841 GAGACCCCTCATCAGGAGCTCAATCCCGGCGAGCGCTTGGAGCTGAGGCGCGGGC 900
      |||
QY 1335 ACAGGGGGTGTGTGATCGACCTTCGACGGAGACGGATGTGACCTCATCTTTGTCCCAT 1394
      |||
Db 901 ACAGGGGGTGTGTGATCGACCTTCGACGGAGACGGATGTGACCTCATCTTTGTCCCAT 960
      |||
QY 1395 GAGAGTCCATGCTCAGCGCTGCTGCTTTCGGGGCAATCAGGCTTCAACACAAC 1454
      |||
Db 961 GAGAGTCCATGCTCAGCGCTGCTGCTTTCGGGGCAATCAGGCTTCAACACAAC 1020
      |||
QY 1455 TGCTCGAGTGTGTGACGACCGCGTTTGGGGCTTTGCCAGGGGAGCTAAGGTGCTG 1514
      |||
Db 1021 TGCTCGAGTGTGTGACGACCGCGTTTGGGGCTTTGCCAGGGGAGCTAAGGTGCTG 1080
      |||
QY 1515 CTCTACACCAAGAGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACCTG 1574
      |||
Db 1081 CTCTACACCAAGAGAGTGGGGCCCACTGAGGATCATCGACGGGGCTCAGGCTACCTG 1140
      |||
QY 1575 TGTGAGATGAGCCCGTGGGACACTTTTGGCTGGGAGAGGATGAAGCAGCATGTGGAG 1634
      |||
Db 1141 TGTGAGATGAGCCCGTGGGACACTTTTGGCTGGGAGAGGATGAAGCAGCATGTGGAG 1200
      |||
QY 1635 GTACGTGGCCAGATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGAGATGAATCA 1694
      |||
Db 1201 GTACGTGGCCAGATGGCAAGATGTGAGCCGGAACGTGGCCAGCGGGAGATGAATCA 1260
      |||
QY 1695 GTCTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACCTGGAG 1754
      |||
Db 1261 GTCTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACCTGGAG 1319
      |||
QY 1755 TGTGGCCAGGATTTCCCGAGCAGGAAATGGCCATTTGATGACACCAATGAATGCATC 1814
      |||
Db 1320 -----GACACCAATGAATGCATC 1337
      |||
QY 1815 CAGTTCCTATTGTTGGCTCGAGACAGCCCGTATGTCTCAACACTATGGAAGCTAC 1874
      |||
Db 1338 CAGTTCCTATTGTTGGCTCGAGACAGCCCGTATGTCTCAACACTATGGAAGCTAC 1397
      |||
QY 1875 AGGTGCGGACCAACAGAAAGTGCAGTTCGGGGCTACGAGCCCAACGAGGATGACAGCC 1934
      |||
Db 1398 AGGTGCGGACCAACAGAAAGTGCAGTTCGGGGCTACGAGCCCAACGAGGATGACAGCC 1457
      |||
QY 1935 TGGGTGGG 1942
      |||
Db 1458 TGGGTGGG 1465
      |||
```

## RESULT 11

```
US-10-037-270-1032
; Sequence 1032, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
```

```
Query Match 51.8%; Score 1345.8; DB 15; Length 1501;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 593 GGAAGACATCTCTGAGCGATGAGTCAACGTGGCCCGTGTGTGCGCCAGCTCTTTTGGCGG 652
      |||
Db 1 GGAAGACATCTCTGAGCGATGAGTCAACGTGTGCTTGTGTGTGCGCCAGCTCTTTTGGCGG 60
      |||
QY 653 ACCTCTGTGGCCCTGTGTGACAGAAAGGCTGTGACGCTACTCTATCTACATTTGCCAA 712
      |||
Db 61 ACCTCTGTGGCCCTGTGTGACAGAAAGGCTGTGACGCTACTCTATCTACATTTGCCAA 120
      |||
QY 713 TTACGGCTCAGGTAATGTGGGGCCCTGATGCCCTCATTTGAAATGACCCCTGAGGCGAGTGA 772
      |||
Db 121 TTACGGCTCAGGTAATGTGGGGCCCTGATGCCCTCATTTGAAATGACCCCTGAGGCGAGTGA 180
      |||
QY 773 CCTCTCCGGGGCAATTCTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGTTCAGCAATA 832
      |||
Db 181 CCTCTCCGGGGCAATTCTGGCGCTCAGAGATGTGGCTGTGAGGCTGGGTTCAGCAATA 240
      |||
QY 833 TACAGGGGGCCGAGGCGTCAAGCTGGGGCCCACTCTCAGCAGAGTGCCTCGGATATCTT 892
      |||
Db 241 TACAGGGGGCCGAGGCGTCAAGCTGGGGCCCACTCTCAGCAGAGTGCCTCGGATATCTT 300
      |||
QY 893 CTGCGACATGAGAAATGGGCTTAACCTTCTTTTCCCAACCCGGGGGATGGCACCTTTGT 952
      |||
Db 301 CTGCGACATGAGAAATGGGCTTAACCTTCTTTTCCCAACCCGGGGGATGGCACCTTTGT 360
      |||
QY 953 GGAAGCTGGGGCCAGTGTGTGTGACAGACCCCAACCCAGCATGGGCGAGGTGTGCGCT 1012
      |||
Db 361 GGAAGCTGGGGCCAGTGTGTGTGACAGACCCCAACCCAGCATGGGCGAGGTGTGCGCT 420
      |||
QY 1013 GGCTGACTTCAACCGGTGATGGCAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCCCA 1072
      |||
Db 421 GGCTGACTTCAACCGGTGATGGCAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCCCA 480
      |||
QY 1073 CCGCCTCTATCTGCAATGAGCACCCATGGGAAGGTCCGCTTCGGGACATCGCCTCACC 1132
      |||
Db 481 CCGCCTCTATCTGCAATGAGCACCCATGGGAAGGTCCGCTTCGGGACATCGCCTCACC 540
      |||
QY 1133 CAAGTTCTCCATGCCCTCCCTGTCCGCAAGGTATCAACCGCGATCTTTGACAAATGACCA 1192
      |||
Db 541 CAAGTTCTCCATGCCCTCCCTGTCCGCAAGGTATCAACCGCGATCTTTGACAAATGACCA 600
      |||
QY 1193 GGAGCTGGAGATCTCTTCAACAACATTTGCTACCGGAGTCTCTCAGCCAAACCGCTCTT 1252
      |||
Db 601 GGAGCTGGAGATCTCTTCAACAACATTTGCTACCGGAGTCTCTCAGCCAAACCGCTCTT 660
      |||
```



```
QY 1253 CCGCGTCATCCGTAGAGACACGGACACCCCTCATCGAGGAGCTCAATCCCGGCGACGC 1312
Db      |||
QY 661 CCGCGTCATCCGTAGAGACACGGACACCCCTCATCGAGGAGCTCAATCCCGGCGACGC 720
Db      |||
QY 1313 CTTGGAGCTGAGGGCCGGGACACAGGGGCTGTGTACCGACTTCGACGAGACGGGAT 1372
Db      |||
QY 721 CTTGGAGCTGAGGGCCGGGACACAGGGGCTGTGTACCGACTTCGACGAGACGGGAT 780
Db      |||
QY 1373 GCTGACCTCATCTTGTGCCATGAGAGATCCATGGCTCAGCCGCTGTCCGCTCTTCCGGGG 1432
Db      |||
QY 781 GCTGACCTCATCTTGTGCCATGAGAGATCCATGGCTCAGCCGCTGTCCGCTCTTCCGGGG 840
Db      |||
QY 1433 CAATCAGGCTTCAACAACAACTGGCTGGAGTGGTGCACACACCCGCTTGGGGCCCTT 1492
Db      |||
QY 841 CAATCAGGCTTCAACAACAACTGGCTGGAGTGGTGCACACACCCGCTTGGGGCCCTT 900
Db      |||
QY 1493 TGCCAGGGGAGTAAAGTCTGTCTCTACACCAAGAGAGTGGGGCCCACTCAGGATCAT 1552
Db      |||
QY 901 TGCCAGGGGAGTAAAGTCTGTCTCTACACCAAGAGAGTGGGGCCCACTCAGGATCAT 960
Db      |||
QY 1553 CGACGGGGCTCAGGCTACCTGTGTAGATGAGGCCCGTGGCACACTTTGGCTCGGGAA 1612
Db      |||
QY 961 CGACGGGGCTCAGGCTACCTGTGTAGATGAGGCCCGTGGCACACTTTGGCTCGGGAA 1020
Db      |||
QY 1613 GGATGAAGCCAGCACTGTGGAGTGCAGCTGGCCAGATGGCAGATGGTGAAGCCGAAAGT 1672
Db      |||
QY 1021 GGATGAAGCCAGCACTGTGGAGTGCAGCTGGCCAGATGGCAGATGGTGAAGCCGAAAGT 1080
Db      |||
QY 1673 GGCCAGGGGAGATGAATCAGTCTGTGAGATCTCTTACCCCGGGATGAGGACACACT 1732
Db      |||
QY 1081 GGCCAGGGGAGATGAATCAGTCTGTGAGATCTCTTACCCCGGGATGAGGACACACT 1140
Db      |||
QY 1733 TCAGAACCCAGCCCACTGGAGTGGCCAAAGATTTCCAGCAGGAAATGGCCATTG 1792
Db      |||
QY 1141 TCAGAACCCAGCCCACTGGAGTGGCCAAAGATTTCCAGCAGGAAATGGCCATTG 1200
Db      |||
QY 1793 CATGACACCAATGAATGCATCCAGTTCCTTCCAGTGGCCAAAGATTTCCAGCAGGAAATGGCCATTG 1852
Db      |||
QY 1201 CATGACACCAATGAATGCATCCAGTTCCTTCCAGTGGCCAAAGATTTCCAGCAGGAAATGGCCATTG 1260
Db      |||
QY 1853 TGTCAACACCTATGGAAGTACAGTGGCCGACCAACAAAGAGTGCAGTCCGGGCTACGA 1912
Db      |||
QY 1261 TGTCAACACCTATGGAAGTACAGTGGCCGACCAACAAAGAGTGCAGTCCGGGCTACGA 1320
Db      |||
QY 1913 GCCCAAGAGATGGCAGCCTGCGTGG 1941
Db      |||
QY 1321 GCCCAAGAGATGGCAGCCTGCGTGG 1349
Db      |||
```

## RESULT 12

```
US-10-117-722-1032
; Sequence 1032, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; TITLE OF INVENTION: Polydeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
```

```
; SEQ ID NO 1032
; LENGTH: 1501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(1468)
US-10-117-722-1032
```

```
Query Match      61.8%; Score 1345.8; DB 16; Length 1501;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 593 GGAAGACATCTCTGACGATAGGTCAACGTGGCCCGTGTGTGGCCAGCCTCTTTGCCGG 652
Db      |||
QY 653 AGCTCTCTGGCTCTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTACATTTGCCAA 712
Db      |||
QY 61 AGCTCTCTGGCTCTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTACATTTGCCAA 120
Db      |||
QY 713 TTACGCTTACGTAATGTGGCCCTGATGCCCTCAATGAAATGGACCTTGAGCCAGTGA 772
Db      |||
QY 121 TTACGCTTACGTAATGTGGCCCTGATGCCCTCAATGAAATGGACCTTGAGCCAGTGA 180
Db      |||
QY 773 CCTCTCCCGGGGCATTTCTGGGCTCAGAGATGTGGCTGCTGAGGCTGGGTGAGCAATA 832
Db      |||
QY 181 CCTCTCCCGGGGCATTTCTGGGCTCAGAGATGTGGCTGCTGAGGCTGGGTGAGCAATA 240
Db      |||
QY 833 TACAGGGGGCCGAGGCTCAGCTGGGGCCCATCTCAGAGAGAGTGCCTCGGATPATCTT 892
Db      |||
QY 241 TACAGGGGGCCGAGGCTCAGCTGGGGCCCATCTCAGAGAGAGTGCCTCGGATPATCTT 300
Db      |||
QY 893 CTGGCAATCAGAAATGGCCCTTACTCTTTTCCAAACCGGGGCGATGSCACTTTGT 952
Db      |||
QY 301 CTGGCAATCAGAAATGGCCCTTACTCTTTTCCAAACCGGGGCGATGSCACTTTGT 360
Db      |||
QY 953 GGAGCTGCGGCGAGTGTGTGTGGACAGACCCACAGCATGGCGAGGTGTCGCCCT 1012
Db      |||
QY 361 GGAGCTGCGGCGAGTGTGTGTGGACAGACCCACAGCATGGCGAGGTGTCGCCCT 420
Db      |||
QY 1013 GGCTGACTTCAACCGTGTATGCAAGTGAATCGTCTATGCAACTGGAATGGCCCCCA 1072
Db      |||
QY 421 GGCTGACTTCAACCGTGTATGCAAGTGAATCGTCTATGCAACTGGAATGGCCCCCA 480
Db      |||
QY 1073 CCGCTCTATCTGCAAAATGAGCACCATGGAAAGTTCGGTTCGGGACATGCGCTACCC 1132
Db      |||
QY 481 CCGCTCTATCTGCAAAATGAGCACCATGGAAAGTTCGGTTCGGGACATGCGCTACCC 540
Db      |||
QY 1133 CAAGTTCTCCATGCCCTCCCTGTCCGCAAGGTGCATCACCCCGACTTTTGACAATGACCA 1192
Db      |||
QY 541 CAAGTTCTCCATGCCCTCCCTGTCCGCAAGGTGCATCACCCCGACTTTTGACAATGACCA 600
Db      |||
QY 1193 GGAGCTGGAGATCTTCTTCAACAATGCTACCGAGTCTCTCAGCACAACCGCTCTT 1252
Db      |||
QY 601 GGAGCTGGAGATCTTCTTCAACAATGCTACCGAGTCTCTCAGCACAACCGCTCTT 660
Db      |||
QY 1253 CCGGCTCATCCGTAGAGACAGGAGACCCCTCATCGAGGAGCTCAATCCCGGAGGC 1312
Db      |||
QY 661 CCGGCTCATCCGTAGAGACAGGAGACCCCTCATCGAGGAGCTCAATCCCGGAGGC 720
Db      |||
QY 1313 CTTGGAGCTGAGGGCCGGGCGACAGGGGCTGTGTGACCGACTTCACCGAGAGCGGAT 1372
Db      |||
QY 721 CTTGGAGCTGAGGGCCGGGCGACAGGGGCTGTGTGACCGACTTCACCGAGAGCGGAT 780
Db      |||
QY 1373 GCTGACCTCATCTTGTGCCATGAGAGTCCATGGCTCAGCCGCTGTCCGCTCTTCCGGGG 1432
Db      |||
QY 781 GCTGACCTCATCTTGTGCCATGAGAGTCCATGGCTCAGCCGCTGTCCGCTCTTCCGGGG 840
Db      |||
QY 1433 CAATCAGGGCTTCAACAACAACTGGCTGGAGTGGTGCACACCGCTTGGGGCCTT 1492
Db      |||
QY 841 CAATCAGGGCTTCAACAACAACTGGCTGGAGTGGTGCACACCGCTTGGGGCCTT 900
Db      |||
```

QY	1493	TGCCAGGGGAGCTAAGGTCTGTCTTACACCAAGAAGAGTGGGGCCCACTGAGGATCAT	1552
DB	901	TGCCAGGGGAGCTAAGGTCTGTCTTACACCAAGAAGAGTGGGGCCCACTGAGGATCAT	960
QY	1553	CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTGGCACACTTTTGGCCTGGGGAA	1612
DB	961	CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTGGCACACTTTTGGCCTGGGGAA	1020
QY	1613	GGATGAAGCCAGCAGCTGTGGAGGTGACGTGGCCAGATGGCAAGATCGTGAGCCGGAAACGT	1672
DB	1021	GGATGAAGCCAGCAGCTGTGGAGGTGACGTGGCCAGATGGCAAGATCGTGAGCCGGAAACGT	1080
QY	1673	GGCCAGGGGGAGATGAACCTCAGTGTGTGGAGATCCTTACCCCCGGGATGAGGACACACT	1732
DB	1081	GGCCAGGGGGAGATGAACCTCAGTGTGTGGAGATCCTTACCCCCGGGATGAGGACACACT	1140
QY	1733	TCAGGACCCAGCCCCCACTGGAGGTGTGCCAAGGATTCTCCAGCAGGAAAATGGCCATTG	1792
DB	1141	TCAGGACCCAGCCCCCACTGGAGGTGTGCCAAGGATTCTCCAGCAGGAAAATGGCCATTG	1200
QY	1793	CATGGACACCAATGAATGCATCCAGTTCCTCCATTGGTGTGCCCTCGAGACAAGCCCGTATG	1852
DB	1201	CATGGACACCAATGAATGCATCCAGTTCCTCCATTGGTGTGCCCTCGAGACAAGCCCGTATG	1260
QY	1853	TGTCACACACCTATGGAAAGCTACAGGTGCCGGACCAACAAGAAGTGCACTCGGGGCTACGA	1912
DB	1261	TGTCACACACCTATGGAAAGCTACAGGTGCCGGACCAACAAGAAGTGCACTCGGGGCTACGA	1320
QY	1913	GCCCAACGAGATGGGCACAGCCTGCGTGG	1941
DB	1321	GCCCAACGAGATGGGCACAGCCTGCGTGG	1349
RESULT 13			
US-09-892-877-18			
; Sequence 18, Application US/09892877			
; Publication No. US20030077809A1			
; GENERAL INFORMATION:			
; APPLICANT: Ruben et. al.			
; TITLE OF INVENTION: 97 Human secreted proteins			
; FILE REFERENCE: P2028P1			
; CURRENT APPLICATION NUMBER: US/09/892,877			
; CURRENT FILING DATE: 2001-06-28			
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658			
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10			
; NUMBER OF SEQ ID NOS: 461			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 18			
; LENGTH: 1143			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: SITE			
; LOCATION: (1100)			
; OTHER INFORMATION: n equals a,t,g, or c			
US-09-892-877-18			

QY	995	TGGGCGAGGTGTCGGCTTGGCTGACTTCAACCGTGAATGCAAAAGTGGCAATCGTCTATGG	1051
DB	188	TGGGGGAGGTGTGCGCCTTGGCTGACATTCAACCGTGAATGCAAAAGTGGCAATCGTCTATGG	247
QY	1055	CAACTGGAAATGGCCCCACCGCCTCTATCTGCAAAATGAGCACCCATTGGGAAGTTCGCGTT	1114
DB	248	CAACTGGAAATGGCCCCACCGCCTCTATCTGCAAGATGAGCACCCATTGGGAAGTTCGCGTT	307
QY	1115	CC-GGGACATCGGC-TCACCCAAAGTTCTCGATGCCCTCCCTCG-TCCGACGCGTCAATCAC	1171
DB	308	CCGGGACATCGGCCTTTCACCCAAAGTTCTCGATGCCCTCCCTCGTTCGACGGTCAATCAC	367
QY	1172	C-GCCGACTTTGACAAATGACCAAGGAGCTGGAG-AUCTTCTTCAACAACTTGTCTACCCG	1229
DB	368	CGGCCGACTTTGACAAATGACCAAGGAGCTGGAGAACTTCTTCAACAACTTGTCTACCCG	427
QY	1230	AGTCTCTACGCCAACCGCCTCTTTCGGGTCATCCGTAGAGAGACACGGAGACCCCTCATC	1289
DB	428	AGTCTCTACGCCAACCGCCTCTTTCGGGTCATCCGTAGAGAGACACGGAGACCCCTCATC	487
QY	1290	GAGGAGCTCAATCCCGGCGACCGCTTTGGAGCCTTGGGGCCGGGCAACAGGGGGTGTGGTG	1349
DB	488	GAGGAGCTCAATCCCGGCGACCGCTTTGGAGCCTTGGGGCCGGGCAACAGGGGGTGTGGTG	547
QY	1350	ACCAGCTTTCGACGGAGACGGGATGCTTGGACCTCATCTTGTCCCATGAGAGTTCATGGCT	1409
DB	548	ACCAGCTTTCGACGGAGACGGGATGCTTGGACCTCATCTTGTCCCATGAGAGTTCATGGCT	607
QY	1410	CAGCGGTGTCCGCTTTCGGGGCAATCAGGGCTTCAACACAACTCGCTGCGAGTGGTG	1469
DB	608	CACCGGTGTCCGCTTTCGGGGCAATCAGGGCTTCAACACAACTCGCTGCGAGTGGTG	667
QY	1470	CCACGCAACCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAG	1529
DB	668	CCACGCAACCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAG	727
QY	1530	AGTGGGGCCCACTGTAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCC	1589
DB	728	AGTGGGGCCCACTGTAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCC	787
QY	1590	GTGGCACATTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGAT	1649
DB	788	GTGGCACATTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGAT	847
QY	1650	GGCAAGATGTTAGCCCGGAAACGTGGCCACCGGGGAGATGAACCTCAGTCTGGAGATCCTC	1708
DB	848	GGCAAGATGTTAGCCCGGAAACGTGGCCACCGGGGAGATGAACCTCAGTCTGGAGATCCTC	907
QY	1710	TACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGATTTC	1768
DB	908	TACCCCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGATTTC	967
QY	1770	TCCCAGCAGAAATATGGCCATTGTCATGGACACCAATGAATGCATCCAGTTCCTTCGTG	1828
DB	968	TCCCAGCAGAAATATGGCCATTGTCATGGACACCAATGAATGCATCCAGTTCCTTCGTG	1027
QY	1830	TGCCCTCGAGCAAGACCCGTTATGTGTCAACCTATTGGAACTACAGGTGCGGACCAAC	1888
DB	1028	TGCCCTCGAGCAAGACCCGTTATGTGTCAACCTATTGGAACTACAGGTGCGGACCAAC	1087
QY	1890	AAGAGTGCAGT-CGGGGCTACGAG-CCCAACGAGGATGGCACAGCCTTCGCTGGGC	1943
DB	1088	AAGAGTGCAGTTCGGGGCTACGAGTCCCAACGAGGATGGCACATACGGGCTTGTC	1143
RESULT 14			
US-03-948-783-18			
; Sequence 18, Application: US/09948783			
; Publication No. US20030100051A1			
; GENERAL INFORMATION:			
; APPLICANT: Ruben et. al.			
; TITLE OF INVENTION: 97 Human secreted proteins			
; FILE REFERENCE: P2028P2			

; APPLICANT: Ruben et. al.  
 ; TITLE OF INVENTION: 97 Human secreted proteins  
 ; FILE REFERENCE: PZ028P2

```
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1100)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-18

Query Match          47.1%; Score 1025.6; DB 10; Length 1143;
Best Local Similarity 98.5%; Pred. No. 1.4e-295;
Matches 1119; Conservative 0; Mismatches 9; Indels 8; Gaps 8;

QY      815  GCGTGGGTCAGCAAAATATACAGGGGGCCGAGGGCTGAGCGTGGGGCCCATCTCCAGCAG 874
Db      9   GCGTGGGTCAGCAAAATATACAGGGGGCCGAGGGCTGAGCGTGGGGCCCATCTCCAGCAG 67

QY      875  CAGTGGCTCGGATATCTTCTGCGACAAATGAGAAATGGGCCTTAACCTCTTTCCACACCG 934
Db      68  CAGTGGCTCGGATATCTTCTGCGACAAATGAGAAATGGGCCTTAACCTCTTTCCACACCG 127

QY      935  GCGCGATGGACACCTTTGTGACGCTGGGCGAGTGTGTTGAGACGCCACCCACAGCA 994
Db      128  GCGCGATGGACACCTTTGTGACGCTGGGCGAGTGTGTTGAGACGCCACCCACAGCA 187

QY      995  TGGCGAGGTGTGCGCCTGCTGACTTCAACCGTGTAGTGGCAAGTGGACATGCTATGG 1054
Db      188  TGGCGAGGTGTGCGCCTGCTGACTTCAACCGTGTAGTGGCAAGTGGACATGCTATGG 247

QY      1055  CAAGTGAATGGCCCCCAGCCCTCTATCTGCAAAATGAGACCCATCGGAAGTCCGCTT 1114
Db      248  CAAGTGAATGGCCCCCAGCCCTCTATCTGCAAAATGAGACCCATCGGAAGTCCGCTT 307

QY      1115  CC-GGGACATCGCG-TCACCAAGTTCTCCATGCCCTCCCTCG-TCCGACGGTCAATC 1171
```

```
Db      308  CCGGGACATCGGCTTCCACCAAGTTCTCCATGCCCTCCCTGTTCCGCAAGTCAATC 367
QY      1172  C-GCGGACTTTGACAAATGACAGGAGCTGGAG-ATCTTCTTCAACAACATTCCTACCGC 1229
Db      368  CGGCCGACTTTGACAAATGACAGGAGCTGGAGATCTTCTTCAACAACATTCCTACCGC 427
QY      1230  AGTCTCTTACGCCAAACCGCCTTTCGCGCTCATCGTAGAGACGAGACCCCTCATC 1289
Db      428  AGTCTCTTACGCCAAACCGCCTTTCGCGCTCATCGTAGAGACGAGACCCCTCATC 487
QY      1290  GAGGAGCTCAATCCCGCGAGCGCTTGGAGCTGAGGCGCGGGGCAAGGGGTGGTG 1349
Db      488  GAGGAGCTCAATCCCGCGAGCGCTTGGAGCTGAGGCGCGGGGCAAGGGGTGGTG 547
QY      1350  ACCGACTTTCAGCGAGACGCGGATGCTGGACCTCATCTTGTCCCATGAGAGTCCATGGCT 1409
Db      548  ACCGACTTTCAGCGAGACGCGGATGCTGGACCTCATCTTGTCCCATGAGAGTCCATGGCT 607
QY      1410  CAGCGCGTGTCCGCTTTCGCGGCAATTCAGGGCTTCAACAACATTCGAGTGGTG 1469
Db      608  CAACCGCTGTCCGCTTTCGCGGCAATTCAGGGCTTCAACAACATTCGAGTGGTG 667
QY      1470  CCAGGACCCCGGTTTGGGGCTTTCGAGGGAGCTTAAGTGTGCTCTACACCAAG 1529
Db      668  CCAGGACCCCGGTTTGGGGCTTTCGAGGGAGCTTAAGTGTGCTCTACACCAAG 727
QY      1530  AGTGGGGCCCGCTGAGGATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGAGGCC 1589
Db      728  AGTGGGGCCCGCTGAGGATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGAGGCC 787
QY      1590  GTGGCACACTTTGCGCTTGGGGAAGGATGAAGCCAGCAGTGTGAGGTGAGTGGCCAGAT 1649
Db      788  GTGGCACACTTTGCGCTTGGGGAAGGATGAAGCCAGCAGTGTGAGGTGAGTGGCCAGAT 847
QY      1650  GGCAAGATGTGAGCGGAACTGGCCAGCGGGGAGATGAACCTCAGTGTGAGATCCTC 1709
Db      848  GGCAAGATGTGAGCGGAACTGGCCAGCGGGGAGATGAACCTCAGTGTGAGATCCTC 907
QY      1710  TACCCCGGATGAGGACACACTTTCAGGACCCAGCCCTGAGTGTGGCAGGATTC 1769
Db      908  TACCCCGGATGAGGACACACTTTCAGGACCCAGCCCTGAGTGTGGCAGGATTC 967
QY      1770  TCCAGCAGGAAATGGCCATTCATGGACACCAATGAATGCCATTCAGTTCCTATCCTG 1829
Db      968  TCCAGCAGGAAATGGCCATTCATGGACACCAATGAATGCCATTCAGTTCCTATCCTG 1027
QY      1830  TGGCTTCGACAAAGCCCGTATGTCAACACCTATGAACTACAGTGGCGGACCAAC 1889
Db      1028  TGGCTTCGACAAAGCCCGTATGTCAACACCTATGAACTACAGTGGCGGACCAAC 1087
QY      1890  AAGAGTGCAGT-CGGGGCTACGAG-CCCACAGGATGGCAGCGCTGGGTGGG 1943
Db      1088  AAGAGTGCAGTTCGGGGCTACGAGTCCCAAGGATGGCAGATGGCATATGGGCTGTG 1143

RESULT 15
US-10-641-484
; Sequence 484, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
;             Jeffrey J. Seilhamer
;             Susan G. Stuart
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
```

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/641,643  
FILING DATE: 14-Aug-2003  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 484:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 789 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SYNORAB01  
CLONE: 192279  
SEQUENCE DESCRIPTION: SEQ ID NO: 484 :

Search completed: August 9, 2004, 23:33:53  
Job time : 995 secs

Query Match 36.2%; Score 788; DB 17; Length 789;  
Best Local Similarity 99.9%; Pred. No. 1.3e-224;  
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1058 CTGGAATGCCCCCAACCGCTCTATCTGCAATGAGCACCCCATGGAAGGTCCGCTTCGG 1117  
Db 1 CTGGAATGCCCCCAACCGCTCTATCTGCAATGAGCACCCCATGGAAGGTCCGCTTCGG 60

QY 1118 GGACATCGCCTCACCACAAATTCCTCATGCTCCCTCCCTGCGCACGGTCAATCACCGCCGA 1177  
Db 61 GGACATCGCCTCACCACAAATTCCTCATGCTCCCTCCCTGCGCACGGTCAATCACCGCCGA 120

QY 1178 CTTTGACATGACCGAGGAGCTGGAGTCTTCTTCAACACATGCTCCAGCTCCCTC 1237  
Db 121 CTTTGACATGACCGAGGAGCTGGAGTCTTCTTCAACACATGCTCCAGCTCCCTC 180

QY 1238 AGCCAAACCGCTCTTCCGGGTGATCGGTAGAGCACGGAGACCCCTCATCGAGGAGCT 1297  
Db 181 AGCCAAACCGCTCTTCCGGGTGATCGGTAGAGCACGGAGACCCCTCATCGAGGAGCT 240

QY 1298 CAATCCCGCGCAGCGCTTGGAGCTGAGGGCCGGGCAAGGGGGTGTGGTGACCGACTT 1357  
Db 241 CAATCCCGCGCAGCGCTTGGAGCTGAGGGCCGGGCAAGGGGGTGTGGTGACCGACTT 300

QY 1358 CGACGGAGACGGAGTGTGGAGCTCATCTTGTCCATGAGAGTCCATGGCTCAGCGCT 1417  
Db 301 CGACGGAGACGGAGTGTGGAGCTCATCTTGTCCATGAGAGTCCATGGCTCAGCGCT 360

QY 1418 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTGTGCCACGAC 1477  
Db 361 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTGTGCCACGAC 420

QY 1478 CCGGTTTGGGGCTTTGCCAGGGGAGCTAAGTGTGCTCTACACCAAGAGTGGGGC 1537  
Db 421 CCGGTTTGGGGCTTTGCCAGGGGAGCTAAGTGTGCTCTACACCAAGAGTGGGGC 480

QY 1538 CCACCTGAGGATCATCGAGGGGGCTCAGGCTACCTGTGTGAGATGAGCCCGTGACCA 1597  
Db 481 CCACCTGAGGATCATCGAGGGGGCTCAGGCTACCTGTGTGAGATGAGCCCGTGACCA 540

QY 1598 CTTTGGCCTGGGGAAGGATGAAGCCAGAGTGTGGAGTGAAGTGGCCAGATGGCAAGAT 1657



2929	ThrSerThrAla		---ThrLysSerThrAlaThrSerValThrProIlePro	2944	Db
193	---TCACGTAGGGTCCACGGGCTGAACCCATGTTCACTGACGTACCAACTCAGTTC	249	QY		
2945	SerSerThrLeuGlyThrGlyThrLeuProGluGlnThrThrPro	2961	Db		
250	TGCCTCTGACTAGACAGTAATCCACCCAGCTCAACT	303	QY		
2962	---ValAlaThrMetSerThrIleHisProSerSerThrProGluThrThrHisThrSer	2980	Db		
304	CTGATGCGACCATGATGGGACATTTGAGATCGTCTGGGGGTACAAATGGACCCCAACC	363	QY		
2981	ThrValLeuThrThrLysAlaThrThrArgAlaThr	3999	Db		
364	TGGTTCTGAAGTACACGGGCCACAGAGCGGTGGTGAACATCGCGGTGATGAGCGCA	423	QY		
3000	---SerThrProGlyThrThrThrIleLeuThrGluLeuThrAla	3014	Db		
424	GCTCACCTACTACGCGTGGCGACCGGAGGGGAACGCGCATTTGGGGTCAACAGCCT	480	QY		
3015	Ala	3032	Db		
481	---GCGACATCGACGGGACGCGGGGAGAGATCT	513	QY		
3033	ThrThrTrpIleLeuThrGluLeuThrThrAlaThrThrAlaThrThrAlaSerThrGlySer	3052	Db		
514	ACTCTCTCAACACCAATAATGCTTCGCGGGGTGG	549	QY		
3053	ThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluProSerThrThr	3072	Db		
550	---CCAGCTACACCGACAGTAGTTTCAAGTTCGCGCAATAACCGGT	591	QY		
3073	AlaThrValThrAlaProProGlySerThrAlaThrAlaSerSerThrGlnAlaThrAla	3092	Db		
592	GGGAAGACATCTGAGCGATGAGGTCAACGTGGCCCGTGGTG	636	QY		
3093	GlyThrProHisValSerThrThrAlaThrThrProThrValThrSerSerLysAlaThr	3112	Db		
637	CGAGCTTTTGGCGGACGCTCTGGCTGTGGACAGAAAGGCTCTGGACGCTACT	696	QY		
3113	ProSerSerProGly	3126	Db		
697	CTATCTACATTGCCAATTACGCTAGCGTAATGTGGGCCCTGATGCCCTCA	747	QY		
3127	ArgSerThrAlaThrProThrAlaThrSerThrPheThrAlaIleProSerSerSerLeu	3146	Db		
748	---TTGAAATGGACC	798	QY		
3147	GlyThrThrTrpThrArgLeuSerGlnThrThrThrProThrAlaThrMetSerThrAla	3166	Db		
799	---GAGATGGCTGTGCTGAGGCTGGGTGAGCAAT	831	QY		
3167	ThrProSerSerThrProGluThrValHisThrSerThrValLeuThrThrAlaThr	3186	Db		
832	ATACAGGGGCGGAGCGCTGACGCGGGCCCATCCTCA	879	QY		
3187	ThrThrGlyAlaThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThr	3206	Db		
880	CCTCGGATATCTTCGACCAATGAGATGGCGCTTAACCTTCTTTTCACAAACCGGGCG	939	QY		
3207	ThrLysValProThrThrThr	3222	Db		
940	ATGGACACCTTTGTGACCGTGGCGGACGCTGGTGGTG	978	QY		
3223	SerSerProGlyThrAlaLeuThrPro	3241	Db		
979	---ACGACCCACACGATGCGGAG	1011	QY		
3242	ThrThrThrProThrThrSerGlySerThrValThrProSerSerIleProGlyThr	3261	Db		
1012	---TGGTGAATTCACCCGCTGATGGCAAGTGGACATCGTCTATGGCA	1056	QY		







```

Db 148 -----ValGlnProSer-----152
QY 676 GAAAGGCTCTGACGCTACTATATACATTGCCAATTAGCGTACGCTAATGTGGGC 735
Db 153 -----SerSerSerProProIleSerThrVal-----Ser 163
QY 736 CTGATCCCTCAATTGAATGACCTTGAGCCAGTGCACCTCTCCCGGGCAATTCTGGCGC 795
Db 164 ValGlnProSer-----167
QY 796 TCAGAGATGTGGCTGTAGGCTGGGTGAGCAAAATATACAGGGGGCGGAGGTGACGC 855
Db 168 -----SerSerSerSerAlaProThrThrSerAlaThrSer 179
QY 856 TGGGCCCATCTTCAGCAGCAGTGCCTCGCATCTCT-----GCAGCA 900
Db 180 ValGlnProSerSerSerProProIleSerThrValSerValGlnThrSer 199
QY 901 ATGAGATGGGCTTAACCTCTTTCCACAACCGGGCGATGGCACCTTTGTGGACGCTG 960
Db 200 SerSerSerSerValProThrThrSerThrSerValGlnProSerSerSer 219
QY 961 CGGCCAGTGTGTGTGGACACCCCCACAGCATGGGCGAGGTGTGCGCTGGCTGACT 1020
Db 220 ValProThrThrSerAlaThrSerValArgSerSerSerSerSerProIlePro 239
QY 1021 TCAACCGTGATGCAAGTGGACATCTCTATGCACTGGAATGGCCCGCCACCGCTCT 1080
Db 240 SerThrThrSerValGlnProSerSerSer-----AlaProThrThrSer 256
QY 1081 ATCTGCAATGAGCAGCCATCGGAAGTCCCTTCGCGGACATCGCTCACCCAGTTCT 1140
Db 257 AlaThrSerValGlnPro-----SerSerSerSerThrProIleProSerThr 273
QY 1141 CCATGCCCTCCCTGTGCGCAGGTGATCACCGCGATTGTGCAATACAGCAGAGCTGG 1200
Db 274 ThrSerValGlnProSerSerSerSerAlaProThrThrSerAlaThrSerValGln 293
QY 1201 AGATCTTCTTCAACAACATTGCCT-----ACCGCAGCTCCTCAGCCACCGCTCT 1251
Db 294 ProSerSerSerSerProProIleSerSerThrIleSerValGlnProSerSer 313
QY 1252 TCCGCTCATCCGTAGAGACAGCGAGC-----CCTCATCGAGGAGCTCAATCCG 1305
Db 314 SerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySerAlaPro 333
QY 1306 GCGAGCGCTTGGAGCCTGAGGCGCGGGCACAGGGGGTGTGTGACCGACTTCGACGAG 1365
Db 334 ThrThr-----SerAlaThrSer 339
QY 1366 ACGGATGCTGACCTCATCTGTCCCATGGAGAGTCCATGCTCAGCGCTGTCGCT 1425
Db 340 ValGlnProSerSerSerSerProProIleSerThrIleSerValGlnProSer 359
QY 1426 TCCGGGCAATCAGGGCTTCAACAACACTGCTGCGAGTGTGGCCACGCCCGGTTG 1485
Db 360 SerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySer 379
QY 1486 GGGCCTTTGCCAGGGAGCTAAGTGTGCTCTACACCAAGAGAGTGGGGCCCACTGA 1545
Db 380 AlaPro-----ThrThrSerAlaThrSerValGlnProSerSer 392
QY 1546 GGATCATCGACG---GGGCTCAGGTACTGTGTGAGTGGAGCCCGTGGCACATTGTG 1602
Db 393 SerSerSerValProThrThrSerAlaThrSerValArgSerSer-----408
QY 1603 GCTGGGGAAGATGAAGCAGCAGTGTGGAGTGCAGTGGCCAGATGCAAGATGGTGA 1662
Db 408 -----408
QY 1663 GCGGGAACGTGGCCAGCGGGAGATGAATCACTAGTGTGGAGATCTCTACCCCGGGATG 1722
Db 1722 -----408

```

```

Db 409 SerSerSerThrProIleProThrThrThrSerValGlnProSerSerSerSerVal 428
QY 1723 AGGACACACTTCCAGGACCCAGCCCACTGGAGTGTGCCAAGGATTCTCCACAGGAAA 1782
Db 429 ProThrThrSerAlaThr-----SerValGlnThrSerSerSerSerThr 444
QY 1783 ATGGCCATTGTCATGGACACCAATGAATGCATCCAGTTCCTCCATTCTCGTGT-----1830
Db 445 ProIleProSerThrThrSerValGlnProSerSerSerSerSerAlaProThrThrSer 464
QY 1831 -----GCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGAAGACTACA---GGTGCC 1881
Db 465 AlaThrSerValGlnProSerSerSerSerProProIleSerThrIleSerVal 484
QY 1882 GGCACCAACAAAGATGCAGTGGGGTACAGAGCCACAGAGGATGGACAGCCTGCGTGG 1941
Db 485 GlnProSerSerSerSerProThrThrThrSerValGlnProSerSer 504
QY 1942 GCTGTGGAGCC 1953
Db 505 SerGlySerAla 508

RESULT 4
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C/Accession: T34434
R/Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A/Description: The sequence of C. elegans cosmid K06A9.
A/Reference number: Z21525
A/Accession: T34434
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2232 <GB>
A/Cross-references: EMBL:U0846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A/Experimental source: Strain Bristol N2; clone K06A9
C/Genetics:
A/Gene: CESP:K06A9.1a
A/Map position: X
A/Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1;

```

```

Alignment Scores:
Pred. No.: 2.9e-06 Length: 2232
Score: 242.00 Matches: 189
Percent Similarity: 31.48% Conservative: 88
Best Local Similarity: 21.48% Mismatches: 302
Query Match: 5.97% Indels: 301
DB: 2 Gaps: 30

```

US-09-914-958B-35 (1-2177) x T34434 (1-2232)

```

QY 3 GGAGCTCGAGCCAGCCGCGGACCGGGCTGGGAGCAAGCAGCGCGCGCGCGC 62
Db 458 GlySerSerThrProSerAlaSer-----SerSerSerAlaGlyThrAlaSerThr 474
QY 63 AGAGCGCGCAGCGCGCGCTTCCACGCCCTTAGCGCGCGCGCGCGAGC---GGG 119
Db 475 IleSerGlySerThrGlySerThrAlaThrIleValProGlySerSerSerValGly 494
QY 120 AGATGGCTCCGAGCGGTACCCCGCATGTCCAGATGTTACCGTCTCTGCTGCTGCTC 179
Db 495 SerSerThrGlnSerAla-SerProSerSerProGlyThr-----507
QY 180 TGGTTCTGCCCATCACTGAGGGGTCCAGCGGGCTGAACCCATGTTCACTGCAGTCACC 239
Db 508 -----MetSerThrValSerGlyProThrGly-----516
QY 240 AACTCAGTCTGCTCTGACTATACATAATCCACCCAGCTCAACTATG-----291
Db 517 -----SerThrValThrValProGlySerSerThrSerProAlaPr 531

```

```

QY 291 ----- 291
Db 531 oSerSerSerProAsnProSerSerProAlaSerThrGlySerThrIleSe 551
QY 292 ----GTGGGAGTTACTGATGGACCATGATGGGACTTTGAGATGCTGTCGGGG 347
Db 551 rGlySerSerSerIleValSerThrValSerGlySerThrValSerGlySerThrGl 571
QY 348 TACAATGGACCAACCTGGTTCTGAAGTATGACC----- 381
Db 571 yThrSerGlnSerThrLeuAlaSerSerThrAlaThrProGlySerSerSerThrValPr 591
QY 382 -----GGGCCCCAAGACGGCTGGTGAAC 404
Db 591 oSerSerSerSerProGlnProSerSerGlnSerProAlaProAsnThrGlySerThrTh 611
QY 405 ATCGCGGTGATGAGCCAGCTCACCTACTACGCGCTGGGGACCGGAGGGAAGCC 464
Db 611 rProSerGlnThrSerSerGlnSerProSerProSerMetAsnProSerSerSerThrPr 631
QY 465 ATTGGGG-----TCACAGCTCGGCAC 485
Db 631 oThrGlySerSerGlnSerThrIleThrProGluGlySerThrAlaSerSerProThrGl 651
QY 486 ATCGACGGGACGGCC----- 501
Db 651 ySerThrGlySerThrPheSerValAlaThrGluValThrSerGlnSerThrValProSe 671
QY 502 -----GGGAGGAGACTACTCTCTCAACACCA-----ATAATGCC 536
Db 671 rGlySerSerLeuGlyThrGlnSerThrAsnSerSerProSerProSerSerLeuSerPr 691
QY 537 TTCTCGGGGTGCCACGTACCGACA-----AGTTGTCAAGTTCCGCCAATAAC 587
Db 691 oSerThrSerGlyMetSerThrLeuThrSerGluProSerProSerSerThrGlnSerSe 711
QY 588 CGGTGGGAACACATCTGAGCGATGAGGTCAACGTGGCGGTGGTGGCCAGCTCTTT 647
Db 711 rGlyAlaGlnSerThrLeuThrThrProSerProAsnProSerGlnSerThrSerLe 731
QY 648 GCGGACGCTCTGTGGCTGTGGACAGAAAGGCTCTGGAGCTACTATCTACATT 707
Db 731 uGluSerSerThrSerGlyAlaThrThrSerSerGlySerAlaGlyThrMetThrSe 751
QY 708 GCCAATTACGCTACGTAATGTGGCCCTGATGCCCTCATTTCAATGACACCTGAGGCC 767
Db 751 rProSerGlnSerSerValGlySerSerGlnGlySer----- 764
QY 768 AGTGACCTCTCCGGGGCATTCGGGCTCA--GAGATGGCTGCTGAGGCTGGGTC 824
Db 765 ----ThrSerProAlaAlaSerThrThrSerGlyGluMet-----ThrSerGlnGlySe 781
QY 825 AGCAAAATACAGGGGGCGAGGCGCTACGCGTGGCC----- 861
Db 781 rThrGlnThrProGlySerSerValSerThrSerAlaAlaIleLeuThrSerThrGlnGl 801
QY 862 -----CCATCTCTACAGCAGCATGCC 881
Db 801 nSerValSerThrAsnSerProGlySerThrValThrArgProSerThrValSerGlySe 821
QY 882 TCGGATATCTTCGCAATGAGAAATGGGCTTAATCTCTTTTCCAAACCGGGCGCAT 941
Db 821 rThrSerSerGlySerThrValThrValGlySerThrGluAlaSerThrSerGlySerSe 841
QY 942 GGCA----- 945
Db 841 rValAlaSerSerProAlaProSerThrSerGlnAsnProAsnProSerThrSerSe 861
QY 946 ----CTTTGTGGACGTGGCGCCAGTCTGCT 974
Db 861 rGlySerSerMetIleThrGlnSerProTyProSerGlnSerThrSerProValGluSe 881
QY 975 GTGGACGACCCCC----- 987

```

```

Db 881 rSerThrThrProSerProGlySerProGlyThrThrLeuThrSerThrSerProSerPr 901
QY 988 -----ACCAGCATGGGAGGTGTCGCCTCGCTGACT----- 1020
Db 901 oSerGlnSerThrThrIleGlySerThrGlnGlySerThrSerProGlyIleSerThrTh 921
QY 1021 -----TCAACCGT 1028
Db 921 rSerGluGluMetThrSerGlnGlySerThrGlnThrProGlySerThrGlySerThrVa 941
QY 1029 GATGGCAAGTGGACATGCTCTATGGCAACTGGAATGCCCCACCGCTCTATCTGCAA 1088
Db 941 lThrGlnProSerThrValSerAspSerThrSerSerGlySerThrValThrVal- 959
QY 1089 ATGAGCACCAATGGGAAGGTCCGCTCCGGGACATCGCTCACCCAAAGT----- 1137
Db 960 -----GlySerThrGluGlySerSerSerProIleProSerThrSerGlnAs 975
QY 1138 -----TCTCCATGCCCTCCCTGTCGGCAGC 1163
Db 975 nThrAsnProSerThrSerSerGlySerSerMetSerThrGlnThrProGlnSerSerGl 995
QY 1164 GTCATCACCCCGACTTTGACAATGACAGGAGCTGGAGATCTCTTCA----- 1212
Db 995 nSerThrSerProValGluSerSerThrSerGlyAlaThrSerSerSerGlySerProGl 1015
QY 1213 -ACAACATTCCTACCGCAGCTCTCTCAGCAACCGCTCT----- 1251
Db 1015 yThrThrLeuThrSerIleSerProSerProSerProSerSerThrIleGlySerSerGl 1035
QY 1252 -----TCCGGTTCATCGGTAGAGACGAGACCGGCTCATC 1289
Db 1035 nGlySerThrSerProValValSerThrIleSerGlnGlySerThrGluThrProGlySe 1055
QY 1290 GAGGAGCTCAATCCCGCGAGCGCTTGGAGCTGGAGCGCGGCGCA----- 1335
Db 1055 rThrGlySerThrValThrLysProSerThrValSerGlySerAlaSerSerGlySerTh 1075
QY 1336 -----CAGGGGTGTGTGACCGACTTCGACGAGACGGGATGTGGACCTCATCTTG 1388
Db 1075 xAlaThrMetGlySerThrGluAlaSerSerThrSerGlyGlySerSerThrSerProAs 1095
QY 1389 TCCCATGGAG-----AGTCCATGGCTCAGCCGCTCTCGTCTTCGGGGCAATCAGGCG 1442
Db 1095 nProSerGlnSerThrSerProSerThrSerGlyAlaThrSerSerProGlySerSerGl 1115
QY 1443 TTCAACAACAACACTGGCTGGAGTGGTCCACGACCGCGTTGGGGCTTTGCCAGGGGA 1502
Db 1115 yThrThrLeuThr-----SerIleSerProSerProSerGl 1127
QY 1503 GCTAAGTGTGTC-----TCTACACAAGAAGAGTGGGCGCCACCTGAGG 1547
Db 1127 nSerThrIleGlySerSerSerGlnGlySerThrSerProValValSerThrThrSerGl 1147
QY 1548 ATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGACCGCTGGCACACTTGGCCTG 1607
Db 1147 yAspMetThrSerGlnGlySerThrGlnIleProGlySerThrGlySerThrValThr-- 1166
QY 1608 GGGAGGATGAAGCCAGCAGTGTGGAGTGAAGTGGCCAGATGGCAAGATGTGTGACCGG 1667
Db 1167 -----GlnProSerThr-----GlySerGlySerThrSerThrSerGl 1179
QY 1668 AACGTGGCCACGGGGAGATGAATCAGTGTGGAGATCCTCT-----ACCCCGG 1718
Db 1179 yGluIleThrSerGlnGlySerThrGlnThrProArgSerSerLeuSerThrSerProAl 1199
QY 1719 GATGAGGACACATTCAG-----ACCA 1742
Db 1199 aIleSerThrSerThrGlnGlnSerValSerThrAsnSerProGlySerThrValThrGl 1219
QY 1743 GCCCCTGAGTGTGGCCAAAGGATTTCTCCAGCAGGAGAAATGGCCAATGTGATGACACC 1802

```

Db 1219 nPro-----SerThrValArgGlySerThrSerSerGlySerThrValThrThrGlySe 1237  
 QY 1803 AATGAATGATCCATGTCCTCCATTCGGTCCGTCGAGACAAAGCCCGCTATGTGTCAACACC 1862  
 Db 1237 rThrGluGlySerThrSerSerSerSerSerSerAlaThrSerLeuSerSerSerPr 1257  
 QY 1863 TATGAAGCTACAGGTGCGGACCAAGAAAGTGCAGTCGGGGCTACGAGCCCAACG 1920  
 Db 1257 oValProSerThrSerGlnSerProAsnProSerThrSerGlySerSerThrProThr 1276  
 RESULT 5  
 T29634  
 hypothetical protein C12D12.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C:Accession: T29634  
 R:Nhan, M.; Hawkins, J.  
 submitted to the EMBL Data Library, March 1996  
 A:Description: The sequence of C. elegans cosmid C12D12.  
 A:Reference number: Z20656  
 A:Accession: T29634  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-825 <NHA>  
 A:Cross-references: EMBL:U51998; PIDN:AA96080.1; GSPDB:GN00028; CBSP:C12D12.1  
 A:Experimental source: strain Bristol N2; clone C12D12  
 C:Genetics:  
 A:Gene: CBSP:C12D12.1  
 A:Map position: X  
 A:Introns: 48/1; 86/3; 137/1; 172/3; 224/3; 253/1; 287/3; 328/2; 454/1; 487/3; 692/1  
 C:Superfamily: Epstein-Barr virus membrane antigen gp350  
 Alignment Scores:  
 Pred. No.: 1.02e-05 Length: 825  
 Score: 231.00 Matches: 119  
 Percent Similarity: 37.45% Conservative: 54  
 Best Local Similarity: 25.76% Mismatches: 154  
 Query Match: 5.70% Indels: 135  
 DB: 2 Gaps: 22  
 US-09-914-958b-35 (1-2177) x T29634 (1-825)  
 QY 57 GCGCGCAGAGCGCGCAGCGCGCCCTCCACGCCCTAGCGCGCGCGCAGAGC 116  
 Db 447 GlyThrGluGlyGlyAlaAsnLysProThrValVal----- 458  
 QY 117 GCGAGATGCTCGAGCGCTGACCCCGGATGTCAGGATGTACCGTTCCTGCTG 176  
 Db 459 -----LeuAlaProProThrThrProGlyProThr----- 468  
 QY 177 CTCTGGTTCTGCGCATCATCTGAGGGGTCCAGCGGGCTGAACCCATGTTCACTGCAGTC 236  
 Db 469 -----PheThrProValThrThrValSerThrMetSerProThrThrValThrVal 486  
 QY 237 ACCAATCAGTTCGGCT---CTGACTATGACAGTAAATCCACCCAGCTCACTATGTT 293  
 Db 487 ProThrThrProThrProValPro---ThrThrAsnThrProProAlaAsnProThrTh 506  
 QY 294 GTGCGAGTACTGATGTGGACATGATGGGACATTGAGATCGTGGCGGGGTACAAT 353  
 Db 506 r-----AlaThrProThrThrValGlyThr-----SerLysGlnThrAsnThrI 521  
 QY 354 GGACCCCAACCTGTTCTGAATGATGACCGGG-----CCGAGAGCGG 395  
 Db 521 eSerPro-----HisLeuSerThrIleThrGlySerIleValThrSerThrPro----- 537  
 QY 396 CTGTGTGAACATCGCGGTGATGAGCGCAGCTACCCCTACTAGCGCTGCGGACCGGAG 455  
 Db 538 -----ThrMetAlaProGlnThrSerAlaSerProThrThr----- 549  
 QY 456 GGGAAACGCCA-----TTGGGGTTCACAGCTCGCATCGACGGGAGCGCGGAG 506  
 Db 550 -----ThrProThrHisThrAlaSerGlnProThrThrThrLysProValValThrTh 568

QY 507 GAGATCT-----ACTTCCTCAACACCAATAATGCTTCTCGGGGTG 548  
 Db rAsnSerValThrProSerThrGlyThrThrValProValProThrThrThrGlySe 588  
 QY 549 GCCAGCTACACGACAGAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCGAGC 608  
 Db 588 rProThrThrGlnThrAlaProValThrLysProThr-----ValProSerSerTh 606  
 QY 609 GATGAGTCAACGTGGCGGTGGTGGCCAGCTCTTTGCGGAGACGCTGTGTGGCTGT 668  
 Db 606 rThrGlnThrAlaProProValThrProThrSerGlnPro-----ProVa 622  
 QY 669 GTGACAGAAAGGCTCTGGAGCT---ACTCTATCTAGATTGCCAATACGCTACGCT 725  
 Db 622 lThrThrThrSerLeuLeuThrThrLeuThrThrProThrValProValThrThrTh 642  
 QY 726 AATGTGGGCTGATGCCCTCATTTGAATGAGACCTGAGGCCAGTGACCTCTCCCGGGG 785  
 Db 642 lVal----- 643  
 QY 786 ATTCTGGCGCTCAGAGATGTGCTGTGAGGTGGGGTGCAGAAATATACAGGGGGCGGA 845  
 Db 643 ----- 643  
 QY 846 GCGTCAGCGTGGGCGCCCATCTCAGCAGAGTGCCTCGGATATCTTCTCGGACAATGAG 905  
 Db 644 -----ProSerSerAlaThrValProThrThrProProThr----- 656  
 QY 906 AATGGGCTTAACCTCTTTTCCAAACCGGGGATGGACACCTTTGTGGACGCTCGCGCC 965  
 Db 657 -----ValThrValAlaAlaThrThrSerLysAlaProValValThrThrSerPr 674  
 QY 966 AGTGTCTGTGTGACGACCCACACGAGTGGCGGAGTTCGCGCTGGCTGACTTCAAC 1025  
 Db 674 oThrLeuAlaProThrSerProThrLysLeuProThrSerProPro-----SerTh 691  
 QY 1026 CGTGATGGGCAAGTGGACATCGTCTATGGCACTGGAAATGGCCCCCACCAGCC 1077  
 Db 691 rVal-----GlyThrSerProThrAlaProAlaAs 701  
 QY 1078 -----TCATCTGCAATGAGCACCCATGGAGAGTTCGCTTCGGGACATCGCT---CA 1130  
 Db 701 nLeuThrThrProThrThrAlaProValAsnProThrSerSerThrThrAlaProThrAl 721  
 QY 1131 CCCAAGTTCCTCCATGCCCTCCCTGTCGCGACCGGTATCATCCGCGCTTGTGACAAATGAC 1190  
 Db 721 aProValAsnProThrSerProThrThrAlaProThrValProProValThrThrTh 741  
 QY 1191 CAGAGCTGGAGATCTTCTTCAACACATTCGCTACCGAGTCTCTCAGCCACCGCTC 1250  
 Db 741 rProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrGlnTh 761  
 QY 1251 TTCGCGCTCA-----TCCGTAGAGACGACGAGACCCCTCATCGAGAGCTC 1298  
 Db 761 rThrProThrProValValThrThrProSerThrIleThrProThrThrArg----- 779  
 QY 1299 AATCCCGCGCAGCGCTTGGAGCTTGAGCGCGGGGACAGGGGTGTGGTACGAGCTTC 1358  
 Db 780 -----ProValThrGlnProAlaSerGlnProAlaThrGlnProAlaThrProThrTh 798  
 QY 1359 GACG 1362  
 Db 798 rThr 799  
 RESULT 6  
 A43932  
 mucin 2 precursor, intestinal - human (fragments)  
 N:Alternate names: mucin SMDC-41  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Mar-1993 #sequence\_revision 12-Apr-1996 #text\_change 05-Nov-1999  
 C:Accession: A49963; A45106; B45106; A43932; B3532; A61257; P00328; P00329  
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994  
 A>Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the  
 A;Reference number: A49963; MUID:94132002; PMID:8300571  
 A;Accession: A49963  
 A;Molecule type: mRNA  
 A;Residues: 1-639 <GU1>  
 A;Cross-references: GB:I21998  
 R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.  
 J. Biol. Chem. 267, 21375-21383, 1992  
 A>Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr  
 A;Reference number: A45106; MUID:93016075; PMID:1400449  
 A;Accession: A45106  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 626-1895 <GU2>  
 A;Cross-references: GB:M94131; NID:gl86395; PIDN:AAA59163.1; PID:gl86396  
 A;Note: sequence extracted from NCBI backbone (NCBIP:116706)  
 A;Accession: B45106  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 2037-3020 <GU3>  
 A;Cross-references: GB:M94132; NID:gl86397; PIDN:AAA59164.1; PID:gl86398  
 A;Experimental source: colon  
 A;Note: sequence extracted from NCBI backbone (NCBIP:116698)  
 R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M  
 J. Clin. Invest. 88, 1005-1013, 1991  
 A>Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp  
 A;Reference number: A43932; MUID:91358717; PMID:1885763  
 A;Accession: A43932  
 A;Molecule type: DNA  
 A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
 A;Cross-references: GB:M74027; NID:gl88863; PIDN:AAA59875.1; PID:gl88864  
 A;Note: sequence inconsistent with the nucleotide translation  
 A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)  
 R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.  
 J. Biol. Chem. 264, 6480-6487, 1989  
 A>Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden  
 A;Reference number: A33532; MUID:89197956; PMID:2703501  
 A;Accession: B33532  
 A;Molecule type: mRNA  
 A;Residues: 1916-2193 <GU4>  
 A;Cross-references: GB:M22405; NID:gl88873; PIDN:AAA36334.1; PID:gl88874  
 A;Experimental source: intestine  
 R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.  
 J. Clin. Invest. 87, 77-82, 1991  
 A>Title: Human bronchus and intestine express the same mucin gene.  
 A;Reference number: A61257; MUID:91086481; PMID:1985113  
 A;Accession: A61257  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>  
 A;Experimental source: bronchus  
 R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,  
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
 A>Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t  
 A;Reference number: PQ0328; MUID:92198477; PMID:1550588  
 A;Accession: PQ0328  
 A;Molecule type: mRNA  
 A;Residues: 2328-2468 <XUG>  
 A;Cross-references: GB:M86523  
 A;Experimental source: small intestine  
 A;Accession: PQ0329  
 A;Molecule type: protein  
 A;Residues: 2328-2342, 'K', 2344-2354 <XUG1>  
 C;Genetics:  
 A;Gene: GDB:MUC2  
 A;Cross-references: GDB:120203; OMIM:158370  
 A;Map position: lip15.5-lip15.5  
 C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von  
 C;Keywords: glycoprotein; intestine; tandem repeat  
 F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>  
 Alignment Scores:

Pred. No.:	1.74e-05	Length:	3020
Score:	229.00	Matches:	162
Percent Similarity:	29.79%	Conservative:	65
Best Local Similarity:	21.26%	Mismatches:	251
Query Match:	5.65%	Indels:	284
DB:	2	Gaps:	36

US-09-914-958b-35 (1-2177) x A43932 (1-3020)

QY	124	TGCTCCGAGCGGTGACCCCGCATCTCCA	---	153
Db	1159	TyrLeuGluGlyCysTyrProArgCysProLysAspArgProIleTyrGluGluAspLeu	1178	
QY	154	-----GGATGTTAC	---	162
Db	1179	LysLysCysValThrAlaAspLysCysGlyCysTyrValGluAspThrHisTyrProPro	1198	
QY	163	-----CGTTCCTGCTGCTGCTCT	---	180
Db	1199	GlyAlaSerValProThrGluThrCysLysSerCysValCysThrAsnSerSerGln	1218	
QY	181	-----GOTTCTGC	---	189
Db	1219	ValValCysArgProGluGluGlyLysIleLeuAsnGlnThrGlnAspGlyAlaPheCys	1238	
QY	190	CCA---TCACTGAGGGGTCCTCCAGCGG	---	231
Db	1239	TyrTrpGluIleCysGlyProAsnGlyThrValGluLysHisPheAsnIleCysSerIle	1258	
QY	232	-----CAGTCACCAACT	---	243
Db	1259	ThrThrArgProSerThrLeuThrThrPheThrThrIleThrLeuProThrThrProThr	1278	
QY	244	CAGTTCTGCTCTGACTATGACAGTATCCACCCAGCTCACTAAGTGTGGCAGTTA	303	
Db	1279	SerPheThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	1295	
QY	304	CTGATGTGGACCATGATGGGACTTGTAGATCGTGTGG	---	342
Db	1296	SerThrThrProLysLeuCysCysLeuTrpSerAspTrpIleAsnGluAspHisProSer	1315	
QY	342	-----	---	342
Db	1316	SerGlySerAspAspGlyAspArgGluProPheAspGlyValCysGlyAlaProGluAsp	1335	
QY	343	-----CGGGGTACATGACACCACTGGTCTGAAGTATGACCGGCCACAGA	---	390
Db	1336	IleGluCysArgSerValLysAspProHisLeuSerLeuGluGlnHisGlyGlnLysVal	1355	
QY	391	-----AGCGGC	---	396
Db	1356	GlnCysAspValSerValGlyPheIleCysLysAsnGluAspGlnPheGlyAsnGlyPro	1375	
QY	397	-----TGGTGAACATCGCGGTGG	---	414
Db	1376	PheGlyLeuCysTyrAspTyrLysIleArgValAsnCysCysTrpProMetAspLysCys	1395	
QY	415	ATGACGGCAGCTCACCTCTACTACGGCTGCGGGACCGGACGGGAAACGCCATTGGGTCA	474	
Db	1396	IleThrThrProSerProThrThr	1407	
QY	475	CAGCTCGGACATCGACGGGACCGCGGAGGAGATCTACTCTCTACACCAATAATG	534	
Db	1408	ProProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	1427	
QY	535	CCTTCTCGGGGTGGCCACGTACACGACAAGTTCTTCAAGTTCCGCAATAACCGGTGG	594	
Db	1428	ThrThrThrThrProPro-----ProThrThrProSerProIleThr-----	1443	
QY	595	AAGACATCTGAGCGATGAGTCAACGTGGCCCGGTGGTGGCGAGCTCTTTGGCGGAC	654	
Db	1444	-----ThrThrThrThrProLeu-----ProThrThrThrProSer	1455	

```
QY 655 GCTCTGTGGCTGTGTGACAGAAAGGGCTCTGGACGGCTACTCTATCTACATTCGCAATT 714
Db 1456 Pro-----ProIleSerThrThrThrProProProThrThrThrProSerProPro 1473
QY 715 ACGCTACGGTAATGTGGCCCTGATGCCCTCAATTGMAATGGACCCCTGAGCCAGTGACC 774
Db 1474 ThrThrThr-----ProSer-----ProProThrThr 1482
QY 775 TCTCCCGGGCAATCTGGCGCTCAGAGTGGTGTGTGAGGCTGGGGTTCAGCAATATA 834
Db 1483 ThrProSerPro-----1486
QY 835 CAGGGGCGGAGGCGTCAGCGTGGGCGCCATCCTCAGCAGCAGTGCCTCGGATATCTTCT 894
Db 1487 -----ProThrThrThrThrThrProProProThrThrThrPro-----Ser 1501
QY 895 GCGCAATGAGATGGGCTTAACCTTCCTTTCCAAACCGGGCGATGGACCTTTGTGG 954
Db 1502 ProProMetThrThrProIleThrProProAlaSerThrThrLeuProProThrThr 1521
QY 955 ACG---CTGGCGCAGTGGTGTGGAGCAGCCACCAGCATGGCGAGGTGTCGCC 1011
Db 1522 ThrProSerProProThrThrThrThrThrThrProProProThrThrThrProSerPro 1541
QY 1012 TGGCTGACTTCAACCGGATGGCAAGTGGACATCGTATGGCAACTGGAATGCCGCC 1071
Db 1542 ---ProThrThrThrProIleThrProProProSerThr-----ThrThrLeuProPro 1558
QY 1072 ACCGCTCTATCTGCAATGAGCACCATGGGAAGTCCGCTTCCGGACATCCCTCAC 1131
Db 1559 ThrThrThr-----ProSerProProProThrThrThrThrThrProPro 1573
QY 1132 CCA---AGTTCTCATGCCCTCCCTGTCCGACGGTCATCACCGCGACTTGAACATG 1188
Db 1574 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1593
QY 1189 ACCAGGAGCTGGAGATCTTCTCAACAACATGTGCTACCGCAGTCTCAGCCAACCGCC 1248
Db 1594 Thr-----ThrProProProThrThrThrProSerProProProThr 1606
QY 1249 TCTTCGCGTCACTCGTGTAGAGACGCGAGACCCCTCATCGAGAGCTCAATCCCGCG 1308
Db 1607 ThrThrThrThrThrProProProProThrThrThrProSerProProThrThrThrProIle 1626
QY 1309 ACGCTTGGAGCTGAGGGCGGGGCACAGGGGTGTGTGACGACTTCGACGAGAGC 1368
Db 1627 ThrPro-----ProThrSerThr-----1632
QY 1369 GGATGTGGACCTCATCTTCTCCATGGAGATCCATGGCTCAGCCGCTGTCGTCTTCC 1428
Db 1633 -----ThrThrLeuProProProThrThrPro-----SerProProProThrThr 1647
QY 1429 GGGCAATCAGGCTTCAACAACACTGGGTGGAGTGGTGCACACGACCCCGTTGGGG 1488
Db 1648 ThrThrThrProProProThrThr-----1556
QY 1489 CTTTGGCAGGGAGTAAGTGTGTCTACACCAAGAGATGGGGCCCACTGAGGA 1548
Db 1657 ProSerPro-----ProThrThrThrThrPro-----SerProProIleThr 1670
QY 1549 TCATCAGCGGGGCTCAGGCTACTGTGTGAGATGGAGCCCGTGGCACACTTTGGCGCTGG 1608
Db 1671 ThrThrThrProProProThrThrThrThrProSerSerProIleThr-----1687
QY 1609 GGAAGATGAACCCAGCAGTGTGGAGTGTGGTGCAGTGGCCAGATGGTGGAGCCGGA 1668
Db 1688 -----1689
QY 1669 ACGTGGCAGCGGGAGATGAATCAGTGTGTGAGATCTCTCTACCCCGGGATGAGGACA 1728
Db 1690 SerProProThrThrThrMetThrThrProSerProProThrThrThrProSerProIle 1709
QY 1729 CACTTCAGGACCCGCCCCACTGGATGTGGCCCAAGGATCTCCCGACGAGGAATGGCC 1788
```

```
Db 1710 ThrThrThrThrThrProSerSerThrThrThrProSerProProProThrThrMetThr 1729
QY 1789 ATTGCATGGACACCAATGAATGCATCCAGTCCCATTCGTGTGCCCTCGACACA-----1842
Db 1730 ThrPro---SerProThrThrThrProSerProProProThrThrThrMetThrThrLeuPro 1748
QY 1843 -----AGCCCGTATGTGTCAACACTATGAAGCTACAGGTGCCGAGCA 1887
Db 1749 ProThrThrThrSerSerProLeuThrThrThrProLeuProProSerIleThrProPro 1768
QY 1888 ACAAGAGTGCAGTGGGGCTACAGCCCAACAGAGATGGCACAGCCTGCG-----TGG 1941
Db 1769 ThrPheSerProPheSerThrThrThrProThrThrProCysValProLeuCysAsnTrp 1788
QY 1942 GCTGGT 1947
Db 1789 ThrGly 1790

RESULT 7
Q0BE3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C>Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BAN>
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; He
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-527;
C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.: 2,64e-05 Length: 660
Score: 223.50 Matches: 162
Percent Similarity: 27.30% Conservative: 28
Best Local Similarity: 23.28% Mismatches: 218
Query Match: 5.41% Indels: 288
DB: 1 Gaps: 34

US-09-914-958B-35 (1-2177) x Q0BE3 (1-660)
QY 1965 TCTTCAACACAGGGCTCCACAGCCACGAGGCTGTGCCATCCT-----CGT 1918
Db 130 SerAlaGlyLeuGlySerArgGlyProArgPro-----HisProAlaPheGlnValGln 147
QY 1917 TGG---GCTCGTAGCCCGAGCTGCATCTTGTGTGGTTCGGCACCTGTAGCTTCCATAGG 1861
Db 148 TrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArg 167
QY 1860 TGTGTACACATACGGGCTTGTCTCGAGGGGCACACAGTGGAACTGGATCATTCATGG 1801
Db 167 -----167
QY 1800 TGTCCATGCAATGGCCATTTTCTCTGTGGGAGAATCCTTGGCCACACTCCAGCTGGGGCTG 1741
Db 168 -----GlyHisProProGlyAla 174
QY 1740 GGTCTGAAGTGTCTCTCATCCCGGGGTAGAGATCTCCAGCACTGATTCATCTCCC 1681
Db 175 GlyGlnArgProSerGlyProThrGlyGlyArg-----ProAlaAlaProGlyAlaPro 192
QY 1680 CGCTGGCCACGTTCCGGCTCACCATCTTGCATCTGCCACGTCACCTCCACTCCACTGCTGG 1621
Db 193 GlyThrProAlaAlaProGlyProGlyGlyAlaAlaVal---ProSerGlyAlaThr 211
```

1620 QY TTTATCTCTCCAGGCAAGTGTGCGACGGGCTCCATCTCACACAGGTAGCTGAGC 1561  
 212 Db ProHisProGluArgGly---SerGlyProAlaAspProAlaAlaAlaArgLeuPro 230  
 1560 QY CCCCTCGATGATCTCAGGTGGGCCCACTCTTCTGTGTGAGACGACGACTTAGCTC 1501  
 231 Db ProGlu---ArgGlnGluProArgLeu 238  
 1500 QY CCTTGGCAAGGCCCAACCGGTGCGTGGCACCACCTCGCAGCCAGCTTGTGTAAGC 1441  
 239 Db Pro---GlnAspLeuAlaAla---Ala 245  
 1440 QY CTGATGTCCTCCGGAAGACGACAGCGGTGAGCCATGACCTCTCATGGGACAGATGA 1381  
 246 Db GlnArgCysPro---AlaGlyProProThrArgSerGly--- 258  
 1380 QY GTTCAGCATCCGCTCCGTCGAAGTCGTGCACACACCCCTGTGCCCC--- 1330  
 259 Db ---AlaAlaAlaGlnArgThrHisArgA-gPro---ProGlyCysProArgSerAla 275  
 1329 QY ---GGCCCTCAGCTCCAGGCGTCGCCGGATTGAGCTCTCGATAGGGGGTCTCCGT 1273  
 276 Db ArgAsnProGlyCysProArgThr--- 283  
 1272 QY GCTCTACGATGACCGGAGAGCGGTGGCTGAGGAGCTGCGGTAGGCAATGTTGT 1213  
 284 Db ---TrpArgArgSerGlyAlaGln--- 291  
 1212 QY TGAAGAAGATCTCCAGCTCTCTGTCTATTGTCAAAGTCGGCGGTGATGACCGTGGCAGAC 1153  
 292 Db ---ArgGlyHisProProGlyAlaGlyGlnArg---ProSerGlyPro 306  
 1152 QY ---GGAGG---GCATGAGAACTTGGGTGAGCGATGTCCTCCGA 1114  
 307 Db ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 326  
 1113 QY ---CCTTCCCATGGGTCTCATTTGCAGATAGAGCGGTGGGGCCAT 1063  
 327 Db GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAla 346  
 1062 QY TCC--- 1060  
 347 Db AspProProAlaAlaAlaArgLeuProGluArgGlnGluProArgLeuProGlnAsp 366  
 1059 QY ---AGTTGCATAGACGATGTCACATTGTCATCCATGAGTGAAGTACGACGAGCGGCACAC 1003  
 367 Db LeuAlaAlaAlaGlnArgCysPro---AlaGlyProProProThr 380  
 1002 QY CTCGCCCATGCTGGTGGGGTCTCCACACACGACCTGGCCGACGCTCCACAAAGGTGC 943  
 381 Db ArgSerGlyAlaAlaAlaGlnArgThrHisArgProProGlyCysProArgSerAla 400  
 942 QY CATGCGCCCGGTGTGGAAGAAAGTAAAGTTAGGCCATTCTCATTTGTCGAGAAATATCCG 883  
 401 Db ArgAsnProGlyCysProArgThrTrpArgArgSerGlyAlaGlnArgGlyHisPro 420  
 882 QY AGGCACTGCTGTGAGGATGGGGCCCAAGCTGACGCTCGGCCCTGTATATTGCTGA 823  
 421 Db Pro--- 421  
 822 QY CCCAGCCTCAGCAGCCACATCTCTGAGGCCAGAAATGCCCGGAGAGGTCACTGGCCT 763  
 422 Db ---ProGlyAlaGlyGlnArgPro 428  
 762 QY CAGGGTCAATTTCAATGAGGCGCATCAGGGCCACATTTACCGTAGGCGTAAATGGCAATGT 703  
 429 Db SerGlyPro--- 431  
 702 QY AGATAGATGATGCTCCAGAGCCCTTTCTGTCCACAGGCGCACAGACGCTCCGGCAAGA 643  
 432 Db ---ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaPro 448

642 QY GGCTGGCCACACACCGGGCCAGTTGACCTCATCTCATCTCAGGATGCTTCCCAACCGGTTAT 583  
 449 Db Gly---ProGlyGlyGlyAlaAlaValProSerGly--- 459  
 582 QY TCGGAACTTGAACTTGTGCGGTAGCGGCCACCCCGAGAGGCAATATTGCTGT 523  
 460 Db ---AlaThrProHisProGluArg--- 466  
 522 QY TGAGGAAGTAGTCTCTCCCGCGCGTCCGTCGATGTCGACGAGGTGACCCCAATGG 463  
 467 Db ---GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProGluArg 483  
 462 QY CTTTCCCTCCCGCGTCCC---GCAGCGCTAGTAGGTGAGCTGCGCTCATCGA 412  
 484 Db GlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProPro 503  
 411 QY CGCGATGTTTCCACGCGCTCTTGGGCCCGGTCTACTTTCAGAACCGAGTTGGTCCAT 352  
 504 Db ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGly--- 520  
 351 QY TGTACCCCGCCAGCAGATCTCAAAGTCCCATCATGTGTCACATCAGTAAGTCCACAC 292  
 521 Db Cys---ProArgSerAlaArgAsnPro--- 528  
 291 QY CATAGTTGAGCTGGGTGGGATTACTCTCATAGTCAGGAGGACAGAACTGAGTTGGTGA 232  
 528 Db --- 528  
 231 QY CAGTGAACATGGTTTCAGCCCTGGGACCCCTCAGTGATGGCAGAAACACGACAGCA 172  
 529 Db ---GlyCysProArgThrTrpArgArgArgSerGlyAla 540  
 171 QY GCAGGAACGGTAACATCTCGACATCGCGGGTCAGCGCTCGGAGCATCTCCCTGCTCT 112  
 541 Db ---GlnArgGlyHisProProProGlyAla 549  
 111 QY CGGCCCCCGCGCTAGGGCGCTGGGAAGCGGCGCTGCTGCGCTCTGCGCCCGCGCGC 52  
 549 Db aGlyGlnArg-ProSerGlyProThrGlyGlyArg---ProAlaAlaProGlyAla 567  
 51 QY CG---CGGCTGCTTCTCCAGCCCGCGTCCCGCGCTGGCC 13  
 567 Db roglyThrProAlaAlaProGlyProGlyGlyAlaAla 580

RESULT 8  
 T34433  
 hypothetical protein K06A9.la - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
 C;Accession: T34433  
 R;Geisel, C.; Gattung, S.  
 submitted to the EMBL Data Library, December 1996  
 A;Description: The sequence of C. elegans cosmid K06A9.  
 A;Reference number: Z21525  
 A;Accession: T34433  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1032 <GUI>  
 A;Cross-references: EMBL:U00846; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9.la  
 A;Experimental source: strain Bristol N2; clone K06A9  
 C;Genetics:  
 A;Gene: CESP:K06A9.la  
 A;Map position: X  
 A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 995/1

Alignment Scores:  
 Pred. No.: 9.58e-05 Length: 1032  
 Score: 214.50 Matches: 164  
 Percent Similarity: 37.93% Conservative: 89  
 Best Local Similarity: 24.59% Mismatches: 222  
 Query Match: 5.29% Indels: 193  
 DB: 2 Gaps: 30





submitted to GenBank, March 1992  
 A:Description: The DNA sequence of equine herpesvirus-1.  
 A:Reference number: A36805  
 A:Accession: H36802  
 A:Molecule type: DNA  
 A:Residues: 1-797 <TEL>  
 A:Cross-references: GB:M86664; NID:g330791; PID:AAB02506.1; PID:g330862  
 A:Telord, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.  
 Virology 189, 304-316, 1992  
 A:Title: The DNA sequence of equine herpesvirus-1.  
 A:Reference number: A41831; MUID:92295566; PMID:1318606  
 A:Contents: annotation; possible protein-coding frames  
 A>Note: neither amino acid nor nucleotide sequence is given  
 C:Genetics:  
 A:Gene: 71  
 C:Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog  
 C:Keywords: glycoprotein; transmembrane protein  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-797/Product: glycoprotein X #status predicted <MAT>  
 F:23-465/Region: serine/threonine-rich  
 F:489-797/Domain: equine herpesvirus 1 glycoprotein homolog <EHG>  
 F:766-790/Domain: transmembrane #status predicted <TMN>  
 F:590/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:  
 Pred. No.: 0.000104 Length: 797  
 Score: 213.50 Matches: 113  
 Percent Similarity: 35.70% Conservative: 68  
 Best Local Similarity: 22.29% Mismatches: 197  
 Query Match: 5.26% Indels: 129  
 DB: 1 Gaps: 15

US-09-914-958B-35 (1-2177) x VGBEX1 (1-797)

QY	403	ACATCGCGTGCATGAGCGCGAGCTCACCTACTACGCGCTGGGACCGGAGGGAACG	462
Db	23	ThrThrThrGluThrThrThrSerSerSerThrSerGlySerGlyGlnSerThr	42
QY	463	CCATTGGGTCACAGCCTCGACATCGACGGGAGCGCGGGAGAGATCTACTCTCA	522
Db	43	SerSerGlyThrThrAsnSerSerSerProThrThrSerProProThrThrSerSer	62
QY	523	ACACCAATATGCTTCCTCGGGGTGGCCACGTCACACGACAAAGTTTCAAGTTCGCA	582
Db	63	SerPro-----ProThrSerThrHisThrSerSerProSerSerThrSerThrGlnSer	80
QY	593	ATAACCGTGGAGACATCTCGACGATAGGTCAAGTCAAGTGGCCCGGTGGCCAGCC	642
Db	81	SerSerThrAlaAlaThrSerSerAlaProSerThr-----Ala	94
QY	643	TCTTTGCGGAGCTCTGCGCTGTGGACAGAAAGGCTCTGACGCTACTCTATCT	702
Db	95	SerSerThrThrSerile---ProThrSerThrSerThrGluThrThrThrPro	113
QY	703	ACATTGCCAATACGCTACGCTAATGTGGCCCTGATGCCCTCATTTGAAATGGACCTG	762
Db	114	ThrAlaSerThrThrThrProThrThrThrAlaAlaProThrThrAlaAlaThrThr	133
QY	763	AGGCAGTGACCTCTCCCGGGGCATCTGGCGCTCAGAGATGTGGCTGTGAGGCTGGG	822
Db	134	ThrAlaValThrThr-----AlaAla	140
QY	823	TCAGCAATATACAGGGGCGGAGCGCTCAGCGTGGGCGCCCATCTCAGCAGCAGTGCCT	882
Db	141	SerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSerThrProThrThrThr	160
QY	883	CGGATATCTTCGCGACATGAGATGGCCCTAACTCTCTTTTCCACACCGGGCGATG	942
Db	161	ThrProThrSerThrThrThrThrAlaThrThrThrValProThrThrAlaSerThr	180
QY	943	GCACCTTTGTGAGCCTCGCGCCAGTGTGGTGGAGACCCCGCCAGCATGGCGAG	1002
Db	181	ThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr	200

## RESULT 10

T45463

membrane glycoprotein [imported] - equine herpesvirus 1

C:Species: equine herpesvirus 1

C:Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 17-Mar-2000

C:Accession: T45463

R:Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.

J. Equine Sci. 7, 79-87, 1996

QY	1003	GTGTGCGCTGCTGCTCAACCGTGATGGCAAGTGACATCGTCTATGCACTGGA	1062
Db	201	ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr	220
QY	1063	ATGGGCCCCCAGCGCTCTATCTGCAATGAGCACCATGCGGAAGCTCCGCTTCCGGACA	1122
Db	221	ThrAlaAlaThrThrSer-----SerAlaThr	229
QY	1123	TCGCTCACCCAGTTCTCCATGCGCTCCCTGTCGCGACGTCATCAGCGCGACTTG	1182
Db	230	ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr	249
QY	1183	ACAATGACCGAGGTGAGATCTCTCAACAACATTGCTACCGCAGCTCTCAGCA	1242
Db	250	Thr-----AlaAlaThrThrThrAlaAlaThrThrThrGlySerPro	263
QY	1243	ACCGCTCTTCCGCGTCATCCGTA---GAGACAGCGAGACCCCTCATCGAGGAGCTCA	1299
Db	264	ThrSerGlySerThrSerThrThrGlyAlaSerThrSerThrProSerAlaSerThrAla	283
QY	1300	ATCCCGGCGAGCCTTGGAGCTGAGCGCGGGGCACAGGGGGTGTGGTGAACGACTTCG	1359
Db	284	ThrSerAlaThrProThrSer-----ThrSer	292
QY	1360	ACGGAGACGGGATGCTGGACCTCATCTTGTCCATGGAGATCCATGGCTCAGCCGCTGT	1419
Db	293	ThrSerAlaAlaAlaThrThrSer-----ThrProThr	303
QY	1420	CGCTCTCCGGGCAATCAGGCTTCAACAACAATGGGTGGAGTGGTCCACGACGCC	1479
Db	304	ProThrSerAlaAlaThrSerAlaGluSerThrThr-----GluAlaPro	318
QY	1480	GGTTTGGGGCTTCCAGGGGAGTAAGTGTCTCTACACCAAGAGAGTGGGGCCC	1539
Db	319	-----ThrSerThr-----Pro	322
QY	1540	ACCTGAGGATCATCGACGGGGCTCAGGTACCTGTGTGAGATGGAGCCGTCGACACT	1599
Db	323	ThrThrAspThrThrThrProSerGluAlaThrThrAlaThrThrSerProGluSerThr	342
QY	1600	TTGGCTGGGAAAGGATGAAGCCAGCAGTGGAGGTGACGTGGCCAGATGGCAAGATGG	1659
Db	343	ThrValSerAla-----	346
QY	1660	TGACCGGAGCAGTGGCCAGCGGGGAGATGAATCAGTGTGTGAGATCTCTACCCCGGG	1719
Db	347	-----SerThrThrSerAla	351
QY	1720	ATGAGCACACTTCAGGACCCAGCCCTGAGTGTGGCCAGGATCTCTCCAGCAGG	1779
Db	352	ThrThrThrAlaPheThrThrGluSerHisThrSer-----	363
QY	1780	AAAATGGCCATTGCATGACACCAATGAATGCATCCAGTTCCTCCATTCGTCCTCGAG	1839
Db	364	-----ProAspSerSerThrGlySerThrSerThrAlaGluPro	376
QY	1840	ACAAGCCCGTATGTCAACACCTATGGAAGTACAGGTGCGGACCAACAGAGAGTGA	1899
Db	377	SerSerThrPheThrLeuThrProSerThrAlaThrProSer-----ThrAspGlnPhe	394
QY	1900	GTCGGGGCTACGCGCCCAACG	1920
Db	395	ThrGlySerSerAlaSerThr	401



QY 1885 CCAACAGAGTGCAGTGGGCTACGAGCCCAACG 1920  
Db 519 AsnThrThrSer-----GlyAsnThrThrProSer 528  
RESULT 11  
T45462  
membrane glycoprotein [imported] - equine herpesvirus 1  
C:Species: equine herpesvirus 1  
C:Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 09-Jun-2000  
C:Accession: T45462  
R:Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.  
J. Equine Sci. 7, 79-87, 1996  
A:Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine  
A:Reference number: Z22973  
A:Accession: T45462  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-866 <KIP>  
A:Cross-References: EMBL:D89733; PIDN:BAA20037.1  
A:Experimental source: strain HH1  
C:Genetics:  
A:Note: ORF1  
C:Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog  
F:558-866/Domain: equine herpesvirus 1 glycoprotein homology <EHG>  
Alignment Scores:  
Pred. No.: 0.000267 Length: 866  
Score: 206.50 Matches: 125  
Percent Similarity: 34.31% Conservative: 74  
Best Local Similarity: 21.55% Mismatches: 233  
Query Match: 5.09% Indels: 148  
DB: 2 Gaps: 18  
US-09-914-958b-35 (1-2177) x T45462 (1-866)  
QY 195 ACTGAGGGTCCAGCGGGTGAACCATGTTCTACGTGAGTCAACCAACTCA----- 245  
Db 35 ThrSerGlySerThrSerSerSerGlyThrThrAsnSerSerSerProThr 54  
QY 246 GTTCTGCTCCTGACTATGACAGTAATCCACCGACTCACTATGTTGGGAGTACT 305  
Db 55 ThrSerProThrThrSerSerSer-ProProThrSerThr----- 68  
QY 306 GATGTGACCATGATGGGACTTTGAGATCGTGTGGGGGTACATGAGCAACCACTG 365  
Db 69 -----HisThrSerSerPro----- 73  
QY 366 GTTCTGAGTATCACCGGGGCCAGAGCGCTGGTGAACATCGCGGTCTGATGAGCGAGC 425  
Db 74 -----SerSerThrSerThrGlnSerSerSer-----ThrAlaAlaThrSerSerAl 90  
QY 426 TCACCTTACTACCGTGGGCGGACCGGAGGAGCGCCATTTGGGTGACAGCTGGAC 485  
Db 90 aProSerThrAlaSerSerThrThrSerThrSerThrSerThrGluThrThr 110  
QY 486 ATCGACGGGCGGCGGAGGAGACTACTCTCTCAACCAATAATGCTTCTCGGG 545  
Db 110 rThrThrProThrAlaSerThrThrThrProThrThrThrThrAlaAlaProThrThrAl 130  
QY 546 GTGGCCACGTACACCGCAAGTTGTTCAAGTTCGCGAATAACCGGTGGGAGACATCTG 605  
Db 130 a-----AlaThrThrThrAlaValThrAlaAlaSerThrAlaAlaSerThrSer-- 147  
QY 606 AGCATGAGGTCAACGTGGGCGGTGGTGGCGACGCTCTTTCGCGACGCTCTGTGGCC 665  
Db 148 ----AlaGluThrThrThrAlaThrAlaThrAlaThrSerThrProThrThrThrPr 166  
QY 666 TGTGTGACACGAAAGGGCTCGGACGCTACTCTATCTACATTCGCAATAGCGCTACGGT 725  
Db 166 o-----ThrSerThrThrThrThrThrThrThrThrValProThrThrAlaSerTh 184  
QY 726 AATGTGGGCCCTGATGCCCTCATTTAAATGAGACCTGAGGCCACGTACCTCTCCGGGGC 785  
Db 785  
184 rThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrTh 204  
786 ATTCTGGCGCTCAGAGATGTGCTGCTGAGGCTGGGGTCCAGCAAATATACAGGGGGCCGA 845  
204 r-----ThrAl 206  
846 GGCGTCAGCGTGGGGCCCATCTCTAGCAGAGTGCCTCGGATATCTTCTGGACACATGAG 905  
206 aAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAl 226  
906 AATGGCCTACTCTCTTTCACAAACCGGGCGATGGCACCTTGTGGACGCTCGGGC 965  
226 aAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrSe 246  
966 AGTGTGTGTGTCACGACCCACGATGGCGAGGTGTCGCTCGCTGCTGCTCAACTCAAC 1025  
246 rAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAl 266  
1026 CGTGATGGCAAGTGGACATCGTCTATGGCACTGAATGGCCCGCCCGCTCTCTATCTG 1085  
266 aThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAl 286  
1086 CAATGACGACCCATGGGAGGTTCGCTTCGCGACATCGCTCACCAAGTTCTCCATG 1145  
286 aThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAl 306  
1146 CGCTCCCTGTCGCGACGCTCATCACCGCGCATTTGACATGACACGAGGTGGAGATC 1205  
306 aThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr-----Al 320  
1206 TTCTTCAACAAATGTCCTACGCGAGCTCTCAGCAACCGCTCTTCCGCTCATCCGT 1265  
320 aAlaThrThrThrAlaAlaThrThrThrGlySerProThrSerGlySerThrThr 340  
1266 A-----GAGACGACGAGACCCCTCATCGAGGAGCTCAATCCCGCGCGCTTGGAGCCT 1322  
340 rGlyAlaSerThrSerThrProSerAlaSerThrAlaThrSerAlaThrProThrSer-- 359  
1323 GAGGGCGGGGCACAGGGGTGTGGTACCGACTTCGACGAGACGCGGATGTGGACCTC 1382  
360 -----ThrSerThrSerAlaAlaThrThrSe 369  
1383 ATCTTGTCCATGAGAGTCCATGGCTCAGCCCTGTCTCGCTTCCGGGCAATCAGGC 1442  
369 r-----ThrProThrProThrSerAlaAlaThrSerAl 380  
1443 TTCAACAAACATGGCTGCGAGTGTGCCACGACCCCGTTGGGGCTTTGCCAGGGGA 1502  
380 aGluSerThrThr-----GluAlaPro----- 387  
1503 GCTAAGTCTGTCTTACACCAAGAGTGGGGCCCATCTGAGGATCATCGACGGGGC 1562  
388 -----ThrSerThr-----ProThrThrAspThrThrProSe 399  
1563 TCAGGTACTCTGTGAGTGGACCGCTGGCACCTTTCGCTGGGGAGGATGAAGC 1622  
399 rGluAlaThrThrAlaThrThrSerProGluSerThrThrValSerAla----- 415  
1623 AGCAGTGTGAGGTGACGTGGCCAGATGTCAGAGTGTGAGCGGAAACGTTGCCAGCGGG 1682  
415 ----- 415  
1683 GAGATGAACACTAGTGTGGAGATCTCTACCCCGGGATGAGGACACACTTCAGACCCA 1742  
416 -----SerThrThrSerAlaThrThrThrAlaPheThrThrGl 428  
1743 GCCCATGAGTGTGCCAAGGATCTCCACGAGCAAAATGSCCATGTCATGCACACC 1802  
428 userHisThrSer----- 433  
1803 AATGAATTCACAGTTCCCATTCGTGTGCTCGTGGCCCTCGAGACAGCCCGTATGTCAACACC 1862  
433 oAspSerSerThrGlySerThrSerThrAlaGluProSerSerThrPheThrLeuThrPr 453





```

Db      401 snHisHis-ThrHisAsnHisHisThrHisAsnGlnTrpLeu----- 414
QY      1272 CACGAGACCCCTCATCGAGAGCTCAATCCGGGACGCGCTTGGAGCTCAGGCGCGG 1331
Db      415 HisGlyAspProLeu-----LeuHisProGlyAspHisProHisArgGlnSerAla 431
QY      1332 -----GSCACAGGGGGTGTGGTGACCGACTTCACAGGAGACGGG 1370
Db      432 AspHisHisHisAsnCysAlaThrGlySerMetAlaThr-ProSer-SerSerThr-- 450
QY      1371 ATGTGGAGCTCATCTGTTCCTCATGAGAGTCCATCGCTCAGCGCTGTCCGCTCTCCGG 1430
Db      451 ----GlnThrSerGlyThrPro-----ProSerLeuThrThrAlaThrThr1 466
QY      1431 GGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCGCGGTGTGGGGCC 1490
Db      466 eThrAlaThrGlySerThrThrAsnProSerSer-----ThrProGlyThrThrPr 483
QY      1491 TTTGCCAGGGGAGTAAGTGTGCTCTACACCAAGAAGAGTGGGCGCCACCTGAGGATC 1550
Db      483 olleProValLeuThrSerThrAlaThrThr-----ProAl 496
QY      1551 ATCGACGGGGCTCAGCGTACC 1572
Db      496 aAlaThrSerSerLysAlaThr 503

RESULT 14
T21460
hypothetical protein ZK945.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21460
R:Wilkinson, J.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z19425
A:Accession: T21460
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-796 <WIL>
A:Cross-references: EMBL:Z48582; PIDN:CAA88469.1; GSPDB:GNO0020; CESP:ZK945.10
A:Experimental source: clone F27E5
A:Gene: CESP:ZK945.10
A:Map position: 2
A:Introns: 62/3; 124/3; 169/1; 216/1; 262/3; 371/1; 581/3; 608/1

Alignment Scores:
Pred. No.: 0.000392 Length: 796
Score: 203.50 Matches: 151
Percent Similarity: 36.97% Conservative: 83
Best Local Similarity: 23.85% Mismatches: 210
Query Match: 5.02% Indels: 189
DB: 2 Gaps: 28

US-09-914-958b-35 (1-2177) x T21460 (1-796)
QY      190 CCATCACTAGG-----GGTCCAGCGGGGTGAACCATCT-----TCACTGCAG 234
Db      253 ProThrLeuArgArgMetLysArgAspAlaGlyAspAsnThrCysAspTyrThrIleGlu 272
QY      235 TCACCAACTCAGTTCTGCCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGGTG 294
Db      273 SerThrSer-----ThrSerThrThrThrThrThrThrThrThrThrThrThr 287
QY      295 TGGCAGTTACTGATGTGGACCATATGGGCACTTTGAGATCTGCTGGCGGGGTACATG 354
Db      288 -----SerThrValThrSerThr-----ThrThr 295
QY      355 GACCAACCTGGTTCTGAAGTATGACGGGGCCAGAGCGGCTGGTGAACATCGGGTCG 414
Db      296 ValProThrSerThrSerThrValThr-----ThrAlaMetSer 308

```

```

QY      415 ATGAGCGAGCTCACCTACTAGCGCTCGGGACCGGACGGGAAACGCAATTGGGGTCA 474
Db      309 ThrSerThrSerThrProSerThr-----SerThrThrIleGluSer 322
QY      475 CAGCCTGCACATCGACGGGACCGCGGAGAGATCTACTTCTCAACACCAATAATG 534
Db      323 ThrSerThrThrPheThrSerThrAlaSer---ThrSerThrSerSerThr----- 338
QY      535 CTTTCTCGGGGTGGCCACGTACACCGCAAGTTGTTCAAGTTCGCCAATAACCGGTGGG 594
Db      339 -----SerThrThrGlnGlnSerSerSerThrIleThrSerSer 351
QY      595 AAGACATCTCTAGCGATAGGTCAACGTGCCCGTGGTGGCCAGCCTCTTTGCCGGAC 654
Db      352 ProSerSer----- 354
QY      655 GCTCTGTGGCCTGTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTACA---TTGCCA 711
Db      355 -----ThrThrLeuSerThrSerIlePro 362
QY      712 ATTACGCTTACGGTAATGTGGGCCCTGATGCCCTCATTTGAAATGGACCTTGAGCCAGTG 771
Db      363 ThrThrThrThr-----ProGluIleThrSerThrLeuSerSerLeu 376
QY      772 ACCTCTCCCGGGCATTTCTGGCCTCAGAGATGGCTGTGAGGCTGGGGTCCAGCAAT 831
Db      376 ----- 376
QY      832 ATACAGGGGCGGAGCGTCAGCGTGGGCCCTCTCAGACGACAGTGCCTCGGATATCT 891
Db      377 -----ProAspAsnAlaIleCysSerTyrLeuAsp 386
QY      892 TCTGCGACAATGAGAAATGGCCCTAACTCTCTTTTCCAAACCGGGGATGGCACCTTTG 951
Db      387 GluThrThrThrSer-----ThrThrPheThrThrThr----- 397
QY      952 TGGACGCTGGCGCAGTGTGGTGTGGACGACCCCGACGAGTGGSCGAGGTGTGCGCC 1011
Db      398 -----MetLeuThrSerThrThrThrGluGluProSerThr 409
QY      1012 TGGCTGACTTCAACCGGTGATGGCAAAAGTGACATCGTCTATGGCACTGAATGGCCCCC 1071
Db      410 SerThrThrThrThrGluValThrSer---ThrSerSerThrValThrThrGluPro 428
QY      1072 ACCGCTCTATCTGCAAAATGAGCACCATGGAGGTCCGCTCCGGGACATCGCCTCAC 1131
Db      429 ThrThrThrLeuThrThr-----SerThrAlaSerThrSerThrGlu 443
QY      1132 CCAAGTTCTCCATGCCCTCCCTCGCGACGGTTCATCACCGCGACTTTTGACAATGACC 1191
Db      444 ProSerThrSerThrValThrThrSerProSerThrSerProValThrSerThrValThr 463
QY      1192 AGGAGCTGGAGATCTCTTCAACA---ACATTCGCTACCGCAGCTCTCAGCCAAACGCC 1248
Db      464 SerSerSerSerSerSerSerThrThrValThrThrProThrSerThrGluSerThrThr 483
QY      1249 TCTTCCGCTCATCCGTA---GAGACACGAGAGACCCCTCATCGAGGAGCTCAATCCCG 1305
Db      484 SerProSerSerThrValThrThrSerThrThrAlaProSerThrSerThrThrGlyPro 503
QY      1306 CGACGCTTGGAGCTTGAGCGCGGGGACACGGGGGTGGTGACCGACTTCACCGGAG 1365
Db      504 SerSerSerSerSerThr-----ProSerSerThrAla 514
QY      1366 ACGGAGTGTGGACCTCATCTCTGTCCATGGAGAGTCCATGGCTTCAGCCGCTGTCGCT 1425
Db      515 SerSerSerValSerSerSerThrThrAlaSerSerThrThrGlnSerThrThrGlnSer 534
QY      1426 TCCGGGGCAATCAGGGCTTCAACAACAACCTGGCTGGAGTGGTGGCCACGACCGGGTTG 1485
Db      535 SerThrThrThrLysSerGluThrThrThr-SerSerAspGly---ThrAsnProAspPh 553
QY      1486 GGGCCTTTGCCAGGGGAGCTAAGTCTGCTCTACACCAAGAAGAT----- 1532

```



```
Db      553 e-TyrPheValGluLysAlaThrThrThrPheTyrAspSerThrSerValAsnLeuThrL 573
QY      1533 --GGGGCCACCTGAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGAGACCGG 1590
Db      573 euAsnSerGlyLeuGlyLeile-----GlyTyrGlnThrSerIleGluCyst 589
QY      1591 TGCACACTTTTGGCTTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATG 1650
Db      589 hrSer-----ProThrSerSerAsnTyrValSerThrThrLysAspG 603
QY      1651 GCAAGATGTGTGACCGGACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATCCTCT 1710
Db      603 lYalaCysPheThrLysSerValSerMetProArgLeuGlyGlyThr-TyrProAlaSer 622
QY      1711 ACC-----CCCGGGATGAGGACACACTTCAGACC-----1740
Db      623 ThrPheValGlyProGlyAsnTyrThrPheArgAlaThrMetThrThrAspAspLysLys 642
QY      1741 -----CAGCCCACTGGAGTGTGGCCAAAGGATTCT 1770
Db      643 ValTyrTyrThrTyrAlaAsnValTyrIleGlnGluTyrSerSerThrThrIleGluSer 662
QY      1771 CCAGCAGGAAATGGCCATTGCGATGACACCAATGAATGCATCCAGTCCCATTCGTGT 1830
Db      663 GluSerSerThrSerAlaValAlaSerSerThrSerThrProSerThrProSerSer 682
QY      1831 GCCCTCGACACAAGCCGATGTGTCAACACCTATGGAAGCTACAGGTCCGCGACCAACA 1890
Db      683 ThrLeuSerThrSerThr---ValThrGluProSerSerThrArgSerSerAspSerThr 701
QY      1891 AGAAGTGCACTGGGCTACGAGCCCAAGAGATGGCACAGGATGGCACAGCTCGGTGGTGGA 1950
Db      702 ThrThrSerAlaGlySerThr-----ThrThrLeuGlnGluSerThrThrThrSer-Gl 719
QY      1951 GCCTGTGTGAAGATAGTGACACACAAGT 1981
Db      719 uGluSerThrThrAspSerSerThrThrThr 729

RESULT 15
JQ0405
hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
N:Alternate names: ORF 1 protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C:Accession: JQ0405
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification of
A:Reference number: S04781; MUID:89364717; PMID:2549377
A:Accession: JQ0405
A:Molecule type: DNA
A:Residues: 1-1106 <SH1>
A:Cross-references: EMBL:X15867
A>Note: this reading frame extends between two stop codons and does not begin with a sta
A>Note: the gene encoding this protein overlaps uvrA gene
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:
Pred. No.: 0.000714 Length: 1106
Score: 199.50 Matches: 202
Percent Similarity: 25.88% Conservative: 26
Best Local Similarity: 22.93% Mismatches: 230
Query Match: 4.83% Indels: 423
DB: 2 Gaps: 49

US-09-914-958B-35 (1-2177) x JQ0405 (1-1106)
QY      1848 CGGGTTCTCTCGAGGGCAGCAATGGGAATCGGATTCATTTCATTGGTCCATGCAAT 1789
Db      14 ArgAlaCys-----AlaThrValTyrGlyArgGly-----23
QY      1788 GGCCATTTTCTGTCT---GGGAGAAATCCTTGGGCCACACTCCAGTGGGGCTGGTCTGAA 1732
```

```
Db      24 GlyAspAlaProGlyArgAlaArgLeuSerValGlySerProThrGlyGluThrProVal 43
QY      1731 GTGTGTCTTCATCCCGGGGTAGAGATCTCCAGCACTGAGTTTCATCTCCCCGCTGGCCA 1672
Db      44 SerCysPro-----46
QY      1671 CGTTCGGGTCAACATCTTGGCCATCTGGCCACGTCTCCACTCCACACTGCTGGTTCATCCT 1612
Db      47 -----CysProProLeuCystrip-----52
QY      1611 TCCCCAGGCCAAAGTGTGCCACGGGCT-----1585
Db      53 -----ProArgAlaGlnGluLeuLeuHisArgLeuLeuArgArg 66
QY      1584 -----CCATCTCACACAGGTAGCTAGCCCCCGTCGATGATCCTCAGGT-----GGG 1537
Db      67 ArgGlyProArgArgArgProArgLeuArgProArgArgAlaLeuGlyArgAlaGly 86
QY      1536 CCCCCTCTCTTGTGTGTAGACGACGACCTTAGCTCCCTGGCAAGGCCCAA---ACC 1480
Db      87 ProHis-----ArgArgProGlyArgProArgAlaGlnProGluGlyArg 101
QY      1479 GGTGTGCTGGCCACCACTCGCAGCCAGTTGTTGTAAGCCCTGATTGCCCGGGAACGCG 1420
Db      102 GlyCysGlnLeuPro-----AlaGlyArgHis 110
QY      1419 ACAGCGGCTGAG-----CCATGGACTCTCCATGGGACA 1387
Db      111 GlyArgValHisGlySerValArgLeuGlyGlnValLeuProGlyLeuArgHisAspLeu 130
QY      1386 AGATGAGGTCCAGCATCCCGTCTCCGTGCA-----1357
Db      131 ArgArgGlyProAlaAla---LeuArgArgValAlaLeuLeuLeuArgProHisValPro 149
QY      1356 -----AGTCGGTCAACACACCCC-----1339
Db      150 GlyProGlyGlyGlnAlaGlyArgGlyLeuHisArgGlyProValProGlyArgValHis 169
QY      1339 -----1339
Db      170 ArgProGluValHisGlnProGlnProAlaLeuHisSerGlyAspHisHisArgAspLeu 189
QY      1338 -----CTGTGCCCCCGCCCTCAGGCTCCAGGCGTCCCGG 1303
Db      190 ArgLeuHisAlaProAlaLeuGlyThrCysArgGlyAlaAlaLeuProAlaValArgArg 209
QY      1302 -----GATTGAGCTCTCCGATGAGGGGCTCTCCGTGCT 1270
Db      210 AlaGlyGluProAlaAspProAlaAlaAspArgGlyProAlaArgGlyAlaAlaArgAla 229
QY      1269 -----CTCTACCGATGACGGGAAGAGCGGTTGGCTGAGGAGTGGCGTAGGCAATGT 1216
Db      230 HisProLeuProGlyAlaArgAlaArgGlyPro-----ArgProGlnGlyArgVal---246
QY      1215 TGTTCGAAGAAGATCTCCAGCTCTCTGCT-----CATTGTCAAAGTCGGGGTGTGACCG 1162
Db      247 -----ArgGlyProValGlnGlyProValHisAlaGlyLeuArgArg-----260
QY      1161 TCGGACAGGGGAGGCGCATGGAGAACTTGGGTGAGCGCATGTCCCGAAGCGGACCTTCC 1102
Db      261 -----ArgGlyArg-----GlyAspArgProAla 268
QY      1101 CATGGGTCTCATTTTGACATAGAGCGCGTGGGGCCATTCCAGTTGCCATAGACATGT 1042
Db      269 LeuGlyProAlaArgAlaGluAlaGlyGlnAlaHisHisArg---ArgArgArgGly 287
QY      1041 CCATTTTCCCATCAGGTTTGAAGTCAGCCAGGCGGACACCTCGCCCATGCTGGTGGGGGT 982
Db      288 ProProArgHisGluGlyGlyHisProProAlaProHis-----ArgLeuGlyGlyAsp 305
QY      981 CGTCCACACAGCACTGGCCCGCAGCGTCCCAAAAGGTGCCATCGCCCGGTTGTGAAAA 922
```

Db 306 ArgProGluAlaGlyArgProArgArgGlyArgValArgGlyArgGlyAlaGlyArg 325  
 QY 921 -----GGAAGTTAGGCCCATCTTCATTTGCG----- 895  
 Db 326 ArgGluGlyGlnGlnGluHisArgGlyValArgArgAlaGlyArgGlyGlnProPro 345  
 QY 895 ----- 895  
 Db 346 ValProLeuValLeuArgGluAlaLeuLeuProGlnArgAlaArgAlaAspArgGlyArg 365  
 QY 894 -----AGAGATATCCGAGG-----CACTGC 874  
 Db 366 AspArgAlaAlaLeuValLeuLeuGlnGlnProValArgArgValProArgValHis--- 384  
 QY 873 TGCTGAGGATGGGGCCCAAGCTGACGCTCGGCCCTGTATATTTCTGACCCACGCT 814  
 Db 385 -----ArgHisArgLeuProPro-----AlaGlyGlyProGlyPro 396  
 QY 813 CAGCAGCCACATCTCTGAGCGCCAGATGCCCGGAGA----- 775  
 Db 397 ArgArgGlnArgArgAlaValProAlaArgGlyArgArgAlaValValAlaArg 416  
 QY 775 ----- 775  
 Db 417 GlnValHisLeuGlyLeuLeuAlaAlaArgAlaArgAlaGlyGlnGlyAspGlyLeu 436  
 QY 774 -----GGTCACATGGCCTCAGGTCCATTTCAATGAGGCGATCAGGGCCACATACCCT 721  
 Db 437 LeuProGlyHisProValGluGlyPro-----AspGlyGlyGlyAlaArg 451  
 QY 720 AGCGCTAATTGGCAATGTAGATAGTAGCTCCAGAGCCCTTTCTGTCCACACAGG--- 664  
 Db 452 ArgArg-----ProAlaArgGln 457  
 QY 663 -----CCACAGAGCGTCGGCAAGAGGCTGGCCACACACC 628  
 Db 458 GlyLeuGlnGlyGlyAspValProGlnProValArgProArgAlaProLeuHisHis 477  
 QY 627 GGGCCACGTGACCTCAT----- 610  
 Db 478 GlyLeuArgGlyArgHisProLeuArgAspAlaGlnAlaArgGlyAspArgValGlyArg 497  
 QY 609 CGCTCAGATGTTCTCCACCGTTATTCCGGAACCTTGACAACTTGTGGTGTACGTGG 550  
 Db 498 ArgProArgAlaLeuArgValValHisAlaGlyAsp----- 509  
 QY 549 CCA-----CCCCGAGAGGCATTATTGGTGTTCAGGAAGTAGA 511  
 Db 510 ProValProGlyValProArgGlyProProGlnProHis-----GlyProGlu 525  
 QY 510 TCTCTCCCGCGCTCCCGTCGATGTCGAGCTGTGACCCCAATGGCGTTCCCTGCC 451  
 Db 526 ArgAlaArgGlyArgProValHisArgGlyArg-----HisProProAla 540  
 QY 450 -----GGTCCCGCAGCGCGTAGT 433  
 Db 541 HisAlaArgGlyHisGlyValLeuLeuGlyAlaAlaAspGlyProGlyAlaAlaAsp 560  
 QY 432 AGG-----GTGAGCTGGCTCATCGACCG 409  
 Db 561 ArgGlyProGlyValGlnGlyAspProGlyProAlaGlyValProAlaGlyArgArgPro 580  
 QY 408 CGATGT---TCACCAAGCGCTCTCGG----- 385  
 Db 581 ArgValProGlnProArgAlaAlaGlyArgHisProLeuArgArgGlyProAlaHis 600  
 QY 384 ----CCCGGTCACTTCA----- 370  
 Db 601 ProProGlyHisThrAspArgLeuArgAlaGlyArgArgProLeuArgProArgAla 620  
 QY 369 -----GAACCGGTTGGGTCCATTGTACCCGCGCAGCATCTCAAGTCCCATCAT 316  
 Db 621 ValHisArgProAlaProAlaGlyGlnProProHisArgAspProProAlaProAla 640

QY 315 GPTCCACATCAGTAACAGTCCACACCATAGTTGAGTGGGTGGGATTACTCTCATAGTCAG 256  
 Db 641 GlyProArgGln-HisProHis-----Ar 648  
 QY 255 GAGGCAGAACTGAGTTGGTGACTGCAGTGAACATGGTTTCAGCCCGCTGGAGCCCTCAG 196  
 Db 648 gArgArg-----AlaArgArgGlyHisAspArgArgGlyGlyLeuAs 662  
 QY 195 TGATGGCCAGAAACACAGAGCAGCAGC-----AGGAACGGTAACATCTCGACATGC 145  
 Db 662 parGlyHisArgPro-SerArgGlyArgValArgArgGlyArgAlaLeuGlyLeuP 682  
 QY 144 CGGGGTGAGCGTCCGA-----GCCATC-----CTCCCGCTCTCGGCCCGCC 103  
 Db 682 roGlyGlySerGlnGlyGluHisAlaValArgHisArgArgLeuProLeuArgProProL 702  
 QY 102 CGCTAGGGCGGTGGGAAGC-----GGCGCTCGCTCGCGCTCGAGCCT 64  
 Db 702 euHisArgGlyAlaGlyAlaAlaSerArgProGlyGluGlyAlaArgAlaAspGlyProA 722  
 QY 63 TGCCGCGCGCGCGCGCTGCTTGTCTCCAGCCCGCTCCCGGCTGGGCTCGAGCCT 5  
 Db 722 rgArgProGly-----GluGlnProGluGlyArgLeuGlyProGlyPro 736

Search completed: August 3, 2004, 19:49:42  
 Job time : 87.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 3, 2004, 19:37:19 ; Search time 23 Seconds  
(without alignments)  
9857.107 Million cell updates/sec

Title: US-09-914-958B-35  
Perfect score: 4056  
Sequence: 1 cggaggagctcgagccagcc.....tcctcttgagaaaaa 2177

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09914958/runat\_03082004\_161651\_29670/app\_query.fasta\_1.2375  
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.0.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914958.cgn 1 24 @runat\_03082004\_161651\_29670 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272.5	6.7	5703	1 MUSB_HUMAN	Q9hc84 homo sapien
2	257	6.3	1367	1 AMYH_YEAST	P08640 saccharomyc
3	253.5	6.2	5179	1 MUC2_HUMAN	Q02817 homo sapien
4	223.5	5.4	660	1 YHL1_EBV	P03181 Epstein-bar
5	213.5	5.3	797	1 VGLX_HSVB	P28968 equine herp
6	203.5	5.0	3178	1 YS89_CABEL	Q93624 caenorhabdi
7	197.5	4.9	907	1 VGP3_EBV	P03200 Epstein-bar
8	197	4.9	1229	1 P121_HUMAN	Q9y2n3 homo sapien
9	192	4.7	662	1 MUC1_XENLA	Q05049 xenopus lae
10	192	4.6	815	1 PYGO_DROME	Q9v9w8 drosophila
11	191.5	4.7	1199	1 P121_RAT	P52591 rattus norv
12	191	4.7	676	1 MUC1_MESAU	Q00528 mesocricetu
13	190.5	4.6	1185	1 DRPL_HUMAN	P54259 homo sapien
14	186.5	4.6	1140	1 YW96_YEAST	Q04893 saccharomyc
15	186	4.6	815	1 PYGO_DROME	Q9v9w8 drosophila
16	184.5	4.5	1460	1 CAL1_CANFA	O9xs17 canis famil
17	184	4.5	1233	1 MUSA_HUMAN	P28008 homo sapien
18	182	4.4	1255	1 MUC1_HUMAN	P15941 h mucin 1 p

19	181.5	4.5	1185	1 DRPL_HUMAN	P54259 homo sapien
20	180	4.4	1595	1 LTBL_HUMAN	Q14766 homo sapien
21	179.5	4.4	1453	1 CALL_CHICK	P02457 gallus gall
22	179.5	4.4	1464	1 CALL_HUMAN	P02452 homo sapien
23	179	4.4	1161	1 DAN4_YEAST	P47179 saccharomyc
24	179	4.3	1509	1 GSRI_HUMAN	Q9nmz4 homo sapien
25	178.5	4.3	830	1 SREK_HUMAN	Q14162 homo sapien
26	178.5	4.4	1418	1 CAL2_HUMAN	P02458 homo sapien
27	178	4.4	784	1 YAV2_XANCV	P14728 xanthomonas
28	177.5	4.4	623	1 CA44_RABIT	P55787 oryctolagus
29	177.5	4.4	1336	1 W146_HUMAN	Q9c0j8 homo sapien
30	176.5	4.4	1446	1 IE18_PRVKA	P33479 pseudorabie
31	175.5	4.2	1487	1 ICP4_HSVK	P17473 equine herp
32	175	4.3	725	1 AGAL_YEAST	P32323 saccharomyc
33	174.5	4.3	1461	1 IE18_PRVTF	P11675 pseudorabie
34	174.5	4.2	2142	1 BAT2_HUMAN	P48634 homo sapien
35	172.5	4.3	704	1 FBL1_CHICK	O73775 gallus gall
36	172.5	4.2	1487	1 ICP4_HSVB	P28925 equine herp
37	171.5	4.2	1048	1 P100_HCMVA	P08318 human cytom
38	171	4.2	1183	1 DRPL_RAT	P54258 rattus norv
39	171	4.2	1240	1 YQ03_CABEL	Q09550 caenorhabdi
40	171	4.2	1690	1 CA44_HUMAN	P53420 homo sapien
41	170.5	4.2	892	1 HIC1_MOUSE	Q9riy5 mus musculu
42	170.5	4.2	1775	1 CA14_DROME	P08120 drosophila
43	170.5	4.2	2090	1 N214_HUMAN	P35658 homo sapien
44	169.5	4.1	1183	1 DRPL_RAT	P54258 rattus norv
45	169.5	4.1	1367	1 AMYH_YEAST	P08640 saccharomyc

## ALIGNMENTS

RESULT 1  
MUSB\_HUMAN  
ID MUSB\_HUMAN STANDARD; PRT; 5703 AA.  
AC Q9HC84; O00447; O00573; O14985; O15494; O95291; O95451; Q14881;  
AC Q99552; Q9UE28;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High  
DE molecular weight salivary mucin Mgl) (Sublingual gland mucin).  
GN MUC5B OR MUC5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 1-1594 FROM N.A.  
RA Chen Y., Di Y.P., Wu R.;  
RT "Molecular cloning of the amino-terminal and 5'-flanking region of the  
RT human MUC5B mucin gene.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-1325 FROM N.A.  
RX MEDLINE=99009274; PubMed=9790959;  
RA Offner G.D., Nunes D.P., Keates A.C., Afshal N.H., Troxler R.F.;  
RT "The amino-terminal sequence of MUC5B contains conserved  
RT multifunctional D domains: implications for tissue-specific mucin  
RT functions.";  
RL Biochem. Biophys. Res. Commun. 251:350-355 (1998).  
RN [3]  
RP SEQUENCE OF 40-1324 FROM N.A.  
RX MEDLINE=99023932; PubMed=9804771;  
RA Desseyn J.-L., Buisine M.P., Porchet N., Aubert J.-P., Laine A.;  
RT "Genomic organization of the human mucin gene MUC5B: cDNA and genomic  
RT sequences upstream of the large central exon.";  
RL J. Biol. Chem. 273:30157-30164 (1998).  
RN [4]  
RP SEQUENCE OF 1326-4895 FROM N.A.  
RX MEDLINE=97166151; PubMed=9013550;  
RA Desseyn J.-L., Guyonnet-Duperrat V., Porchet N., Aubert J.-P.,  
RA Laine A.;

RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes  
RT various alternate subdomains resulting in a super-repeat. Structural  
RT evidence for a 11p15.5 gene family.";  
RL J. Biol. Chem. 272:3168-3178(1997).  
RN [5]  
RP SEQUENCE OF 4057-4480 FROM N.A.  
RC TISSUE=Salivary gland;  
RX MEDLINE=97292340; PubMed=9147051;  
RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,  
RA Hannibal J., Clausen H.;  
RT "Identification of a major human high molecular weight salivary mucin  
RT (MG1) as tracheobronchial mucin MUC5B.";  
RL Glycobiology 7:413-419(1997).  
RN [6]  
RP SEQUENCE OF 4721-5703 FROM N.A.  
RC TISSUE=Gall bladder;  
RX MEDLINE=97293229; PubMed=9164870;  
RA Keates A.C., Nunes D.P., Afzal N.H., Troxler R.F., Offner G.D.;  
RT "Molecular cloning of a major human gall bladder mucin: complete C-  
RT terminal sequence and genomic organization of MUC5B.";  
RL Biochem. J. 324:295-303(1997).  
RN [7]  
RP SEQUENCE OF 4809-5687 FROM N.A.  
RC TISSUE=Sublingual gland;  
RX MEDLINE=96125355; PubMed=8554565;  
RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;  
RT "Molecular cloning of a novel high molecular weight mucin (MG1)  
RT from human sublingual gland.";  
RL Biochem. Biophys. Res. Commun. 217:1112-1119(1995).  
RN [8]  
RP SEQUENCE OF 4859-5703 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=97347489; PubMed=9201995;  
RA Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.;  
RT "Genomic organization of the 3 region of the human MUC5B mucin.";  
RL J. Biol. Chem. 272:16873-16883(1997).  
CC 1- FUNCTION: Salivary mucin that is thought to contribute to the  
CC lubricating and viscoelastic properties of whole saliva.  
CC 1- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also  
CC in submaxillary glands, endocervix, gall bladder, and pancreas.  
CC 1- PTM: Highly glycosylated.  
CC 1- SUBCELLULAR LOCATION: Secreted.  
CC 1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
CC 1- SIMILARITY: Contains 3 WFPC domains.  
CC 1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF107890; AAC33673.1; -;  
DR EMBL; AF086604; AAC67545.1; -;  
DR EMBL; AJ004862; CAA06167.1; -;  
DR EMBL; Z72496; CAA96577.1; -;  
DR EMBL; X74955; CAA52910.1; -;  
DR EMBL; U63836; AAB61398.1; -;  
DR EMBL; U78554; AAC51344.1; -;  
DR EMBL; U78552; AAC51344.1; JOINED.  
DR EMBL; U78553; AAC51344.1; JOINED.  
DR EMBL; U78551; AAB61343.1; -;  
DR EMBL; U95031; AAB65151.1; -;  
DR EMBL; Y09788; CAA70926.1; -;  
DR Genew; HGNC:7516; MUC5B.  
DR MIM; 600770; -;  
DR GO; GO:0005515; F:protein binding; IPI.  
DR InterPro; IPR006208; Cys\_knot.  
DR InterPro; IPR006207; Cys\_knot\_C.  
DR InterPro; IPR009041; PMP\_inhibitor.

DR InterPro; IPR002919; TIL\_Cysrich.  
DR InterPro; IPR006552; VC\_Out.  
DR InterPro; IPR001007; VWF\_C.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF01826; TIL; 1.  
DR Pfam; PF00093; VWF; 1.  
DR Pfam; PF00094; vwd; 4.  
DR SMART; SM00214; VMC; 6.  
DR SMART; SM00215; VMC\_out; 4.  
DR SMART; SM00216; VWD; 4.  
DR PROSITE; PS01185; CTCK\_1; 1.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS01208; WFPC\_1; 2.  
DR PROSITE; PS0184; WFPC\_2; 2.  
KW Glycoprotein; Repeat; Signal; Polymorphism.  
FT SIGNAL 1 25  
FT CHAIN 26 5703  
FT DOMAIN 77 225  
FT DOMAIN 329 385  
FT DOMAIN 426 580  
FT DOMAIN 858 913  
FT DOMAIN 896 1044  
FT DOMAIN 1457 1603  
FT DOMAIN 1609 4873  
FT DOMAIN 5005 5178  
FT DOMAIN 5353 5425  
FT DOMAIN 5462 5528  
FT DOMAIN 5594 5683  
FT DISULFID 5594 5645  
FT DISULFID 5622 5663  
FT DISULFID 5626 5675  
FT DISULFID 5643 5678  
FT DISULFID 5645 5682  
FT CARBOHYD 145 145  
FT CARBOHYD 201 201  
FT CARBOHYD 254 254  
FT CARBOHYD 402 402  
FT CARBOHYD 516 516  
FT CARBOHYD 806 806  
FT CARBOHYD 930 930  
FT CARBOHYD 1277 1277  
FT CARBOHYD 1293 1293  
FT CARBOHYD 1557 1557  
FT CARBOHYD 1775 1775  
FT CARBOHYD 2192 2192  
FT CARBOHYD 2721 2721  
FT CARBOHYD 3419 3419  
FT CARBOHYD 3948 3948  
FT CARBOHYD 4745 4745  
FT CARBOHYD 4901 4901  
FT CARBOHYD 4958 4958  
FT CARBOHYD 4965 4965  
FT CARBOHYD 4987 4987  
FT CARBOHYD 5037 5037  
FT CARBOHYD 5052 5052  
FT CARBOHYD 5156 5156  
FT CARBOHYD 5427 5427  
FT CARBOHYD 5467 5467  
FT CARBOHYD 5506 5506  
FT CARBOHYD 5507 5507  
FT CARBOHYD 5543 5543  
FT CARBOHYD 5553 5553  
FT CARBOHYD 5604 5604  
FT CARBOHYD 5618 5618  
FT CARBOHYD 5662 5662  
FT VARIANT 5137  
FT CONFLICT 34  
FT CONFLICT 95  
FT CONFLICT 104  
FT CONFLICT 142  
FT CONFLICT 225  
G -> E (IN REF. 2).  
FPGLCN -> LPCLCK (IN REF. 2).  
S -> C (IN REF. 2).  
E -> K (IN REF. 1).  
R -> S (IN REF. 2).  
T -> S (in dbSNP:2672788).  
/FTID=VAR\_014123.

FT CONFLICT 330 331 PL -> T (IN REF. 2).  
FT CONFLICT 337 337 E -> N (IN REF. 2).  
FT CONFLICT 356 356 E -> K (IN REF. 2).  
FT CONFLICT 362 362 G -> R (IN REF. 2).  
FT CONFLICT 369 369 MISSING (IN REF. 2 AND 3).  
FT CONFLICT 374 374 D -> N (IN REF. 2).  
FT CONFLICT 393 394 RT -> TR (IN REF. 2).  
FT CONFLICT 468 468 RK -> GR (IN REF. 2).  
FT CONFLICT 512 512 L -> P (IN REF. 2).  
FT CONFLICT 585 587 GAA -> AH (IN REF. 3).  
FT CONFLICT 601 601 A -> S (IN REF. 3).  
FT CONFLICT 628 629 DP -> RS (IN REF. 2).  
FT CONFLICT 633 633 F -> L (IN REF. 2).  
FT CONFLICT 676 676 A -> P (IN REF. 3).

Alignment Scores:  
Pred. No.: 2,14e-07 Length: 5703  
Score: 272.50 Matches: 185  
Percent Similarity: 34.41% Conservative: 91  
Best Local Similarity: 23.07% Mismatches: 269  
Query Match: 6.72% Indels: 257  
DB: 1 Gaps: 39

US-09-914-958b-35 (1-2177) x MUSB\_HUMAN (1-5703)

QY 19 CCCGGACCGGGCTGGAGCAAGCAGCGCGCGCGCGCGCGAGGCGGCGAGC 78  
Db 4215 ProGlyThrAlaProProValLeuThrSerValLeuThrSerProAlaThrThrProAlaThr 4234  
QY 79 GCCCGTCCACGCCCTAGCGCGCGCGCGCGCGCGAGGCGGAGATGGTCCGAGCGCTG 138  
Db 4235 SerSerLysAlaThrSerSerSerProArgThrAlaThrThrLeuProValLeu 4253  
QY 139 ACCCGCGCATGCTCAGGATGTCAGTCTCGTCTCGTCTGTTCTGCGCA 192  
Db 4254 ThrSerThrAlaThrSerThrAlaThrSerValThrProIlePro 4269  
QY 193 ---TCACTAGGGGTCCACGCGGCTGAACCCAGTTTCATGCGAGTCAACCACTCAGTTC 249  
Db 4270 SerSerThrLeuGlyThrGlyThrLeuProGluGlnThrThrPro 4286  
QY 250 TGCTCTGACTATGACAGTAATCCACCCAGCTCACT---ATGGTGGCAGTTA 303  
Db 4287 ---ValAlaThrMetSerThrIleHisProSerThrThrProGluThrThrHisThrSer 4305  
QY 304 CTGATGTGGACCATGATGGGACTTTGAGATCGTGTGGGGGTACAATGGACCCCAACC 363  
Db 4306 ThrValLeuThrThrLysAlaThrThrArgAlaThrSerSerThrThrProSer 4324  
QY 364 TGGTTCTGAAGTATGACGGGCCGAGAGCGGCTGTGTGAATCGCGGTGATGAGCGCA 423  
Db 4325 ---SerThrProGlyThrThrTrpIleLeuThrGluLeuThrThrAla 4339  
QY 424 GCTCACCCCTACTAGCGCTCGGGGACCGCGGAGCGGAGCCATTTGGGGTTCACAGCTT 480  
Db 4340 Ala-----ThrThrAlaGlyThrGlyProThrAlaThrProSerThrProGly 4357  
QY 481 -----GCGACATCGACGGGAGCGGCGGCGGAGGAGATCT 513  
Db 4358 ThrThrTrpIleLeuThrGluLeuThrThrThrAlaThrThrThrAlaSerThrGlySer 4377  
QY 514 ACTTCCTCAACCAATAATGCTTCGCGGGGGTGG-----CCACGTACA 558  
Db 4378 ThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluProSerThrThr 4397  
QY 559 -----CCGACA-----AGTTGTTCAAGTTCGCGAATAACCGGT 591  
Db 4398 AlaThrValThrValProThrGlySerThrAlaThrAlaSerThrGlnAlaThrAla 4417  
QY 592 GGGGAAGACATCCTGAGCGATGAGTCAACGGGCGCGGTG-----TGG 636  
Db 4418 GlyThrProHisValSerThrThrAlaThrThrProThrValThrSerSerLysAlaThr 4437

QY 637 CAGACCTCTTTGCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGCTCTGGACGCTACT 696  
Db 4438 ProSerSerSerProGly-----ThrAlaThrAlaLeuProAlaLeu 4451  
QY 697 CTATCTACATGTCGAATACGCTACGGTAATGTGGCCCTGATGCCCTCA----- 747  
Db 4452 ArgSerThrAlaThrThrProThrAlaThrSerPheThrAlaIleProSerSerSerLeu 4471  
QY 748 ---TTGAATGGACC---CTGAGCGCAGTGCCTCTCCGGGGCATCTCGCGCTCA--- 798  
Db 4472 GlyThrThrTrpThrArgLeuSerGlnThrThrThrProThrAlaThrMetSerThrAla 4491  
QY 799 -----GAGATGGCTGCTGAGGCTGGGGTGCAGCAAT 831  
Db 4492 ThrProSerSerThrProGluThrValHisThrSerThrValLeuThrThrThrAlaThr 4511  
QY 832 ATACAGGGCGGAGGCGTCAGCGTGGGCCCATCTCA-----GAGCAGGTG 879  
Db 4512 ThrThrGlyAlaThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThr 4531  
QY 880 CCTCGGATATCTTCGACACATGAGAAATGGGCTAACTTCCTTTTCCACACCGGGCGG 939  
Db 4532 ThrLysValProThrThrThr-----ThrThrGlyPheThrAlaThrProSer 4547  
QY 940 ATGCACCTTTGGAGCGTGGCGCCAGTCTGGTGG----- 978  
Db 4548 SerSerProGlyThrAlaLeuThrProPro---ValTrpIleSerThrThrThrPro 4566  
QY 979 -----ACGACCCCGCAGCATGGGCGAG---GTGTCGCC 1011  
Db 4567 ThrThrThrThrProThrThrSerGlySerThrValThrProSerSerIleProGlyThr 4586  
QY 1012 -----TGGCTGACTTCAACCGTGCATGCGAAAGTGCAGCATGCTCTATGGCA 1056  
Db 4587 ThrHisThrAlaArgValLeuThrThrThrThrValAlaThrGlySerMetAla 4606  
QY 1057 ACT-----GGAATGGCCCCACCGCTCT 1080  
Db 4607 ThrProSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThr 4626  
QY 1081 ATCTGCAATGAGCACCACCGTGGGAAGTCCGCTTCCGGGACATCGCTCACCCAGTTCT 1140  
Db 4627 ThrIle-----ThrAlaThrGlySerThrThrAsnProSerSer 4639  
QY 1141 -----CCATGCCCTCC---CTGTCCGACGGTCAACACCGCGGACTTTG 1182  
Db 4640 ThrProGlyThrThrProIleProProValLeuThrSerMetAlaThrThrProAlaAla 4659  
QY 1183 ACAATGACGAGGCTGGAGATCTTCTTCAACACATGTGCTACCGCAGCTCTCAGCCA 1242  
Db 4660 ThrSerSerLysAlaThrSerSerSerSerProArgThrAlaThrThrLeuProValLeu 4679  
QY 1243 ACCGCTCTTCCCGCTCATCGGTAGACGACGAGACCCCTCATCGAGGAGCTCAATC 1302  
Db 4680 ThrSerThrAlaThrLysSerThrAlaThrSerPheThrProIleProSerSerThrLeu 4699  
QY 1303 CCGCGCAGCGCTTGGAGCTTGGCGCGGGGACAGGGGGTGTGGTACCGACTTCGACG 1362  
Db 4700 ---TrpThrThrTrpThrValProAlaGlnThrThr-----ProMetSerThr 4715  
QY 1363 GAGACGGGATGCTGGACCTCATCTGTGCC-----ATG 1395  
Db 4716 MetSerThrIleHisThrSerSerThrProGluThrThrHisThrSerThrValLeuThr 4735  
QY 1396 GAGATCCATGGCTCAGCGCTGTCCGTCTTCGGGGCAATCAGGGCTTCAACAACAAC 1455  
Db 4736 ThrThrAlaThrMetThrArgAlaThrAsnSerThrAlaThrProSerSerThrLeu--- 4754  
QY 1456 GGCTGCGAGTGGTCCACGACCGCGGTTTGGGCGCTTTGGCGGCTTTCAGGGGAGCTAAGTCTGTC 1515  
Db 4755 -----GlyThrThrArgIleLeuThrGluLeuThrThrThr 4766  
QY 1516 TCTACACCAAGAGAGTGGGGCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGT 1575





Qy	1266	AGAGAGCAGGAGACCCCTCATCTAGGAGACTCAATCCCGCGACGCTGGAGCCTGAG	1332
Db	695	lThrSerSerThrThrGluSerSerAlaProValPro---ThrProSerSerSerTh	714
Qy	1326	GGCCGGGGACACGGGGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACACTCATC	1395
Db	714	rThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSe	734
Qy	1386	TTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCOGTCTTCCGGGGCAATCAGGGCTTC	1449
Db	734	rAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSe	754
Qy	1446	AACAACAACCTGGCTCGAGGTGGCCACGACCCGGTTTGGGGCTTTGCCAGGGGAGCT	1505
Db	754	rSerThrThr-----GluSerSerSerAlaProValProThrProSerSerSerThrTh	772
Qy	1506	AAGTCTGTCTACACCAAGAAGAGTGGGGCCACCTGAGGATCATCAGCGG-----	1559
Db	772	rGluSerSerSerAlaPro-----ValProThrProSerSerSerThrThrGluSe	789
Qy	1559	-----	1559
Db	789	rSerSerAlaProValProThrProSerSerSerThrThrGluSerSerValAlaProVa	809
Qy	1560	-----GGCTCAGCTACCTGTGTGAGATGAGCCCGTGGCACACTTTGGCT	1606
Db	809	lProThrProSerSerSerSerAsnIleThrSerSerAlaProSerSerThrProPheSe	829
Qy	1607	GGGAGGATGAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCGC	1666
Db	829	rSerSerThrGluSerSerSerValProValProThrProSerSerSerThrThrGluSe	849
Qy	1667	GAACCTGGCC-----AGCGGGGAGATGAATCACTGCTGGAGATCCTCTACCC	1714
Db	849	rSerSerAlaProValSerSerSerThrThrGluSerSerValAlaProValProThrPr	869
Qy	1715	CGGGATGAGGACACACTTCAGGACCAGCCCACTGAGGTGGCCAGGATTCCTCCA	1774
Db	869	oSerSerSerSerAsnIleThrSerSerAlaProSer-SerIleProPheSerSerThr	889
Qy	1775	GCAGGAATGGCCATTGCATGGACACCAATGAATGCATCCAGTCCCATTCGTTGCC	1834
Db	889	hrGluSerPheSerThrGly---ThrThrValThrProSerSerSerSerThrThrProGlyS	908
Qy	1835	TCGAGACAAAGCCCTGTGTCAACACTATGGAAGCTACAGTGCCTGGACCAACAGAA	1899
Db	908	erGlnThrGluThrSerValSerSerThrThrGluThrThr---IleValProThrLysT	927
Qy	1895	GTGCAGTCTGGGGCTACGAGGCCAACG	1920
Db	927	hrThrThrSerValThrThrProSer	935
RESULT 3			
MUC2 HUMAN			
ID	MUC2 HUMAN	STANDARD;	PRT; 5179 AA.
AC	Q02817; Q14878;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Mucin 2 precursor (intestinal mucin 2).		
GN	MUC2 OR SMUC.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_taxid=9606;		
FN	[1]		
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Intestine;		
EX	MEDLINE=94132002; PubMed=8300571;		
EA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;		
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.		
RT	Identification of the amino terminus and overall sequence similarity		
RT	to pig-von Willebrand factor."		

RL J. Biol. Chem. 269:2440-2446(1994).  
 RN [2]  
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
 RC TISSUE-Colon;  
 RA MEDLINE=93016075; PubMed=1400449;  
 RX Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
 RA Kim Y.S.;  
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
 RT both upstream and downstream of its central repetitive region.";  
 RL J. Biol. Chem. 267:21375-21383(1992).  
 RN [3]  
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
 RX MEDLINE=91358717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.M., Kim Y.S.;  
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 RT and polymorphism.";  
 RT J. Clin. Invest. 88:1005-1013(1991).  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and  
 CC other mucus membrane-containing organs. Thought to provide a  
 CC protective, lubricating barrier against particles and infectious  
 CC agents at mucosal surfaces.  
 CC -!- SUBUNIT: Multimeric.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
 CC BRONCHUS, CERVIX AND GALL BLADDER.  
 CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and  
 CC varies among different alleles.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 2 VWFC domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L21998; AAA95295.1; -;  
 DR EMBL; M74027; AAA59875.1; -;  
 DR EMBL; M94131; AAA59163.1; -;  
 DR EMBL; M94132; AAA59164.1; -;  
 DR PIR; A49963; A43932.  
 DR Genew; HGNC:7512; MUC2.  
 DR MIM; 158370; -;  
 DR GO; GO:0005803; C:secretory vesicle; TAS.  
 DR InterPro; IPR006208; Cys\_knot.  
 DR InterPro; IPR006207; Cys\_knot\_C.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR002919; TIL\_Cysrich.  
 DR InterPro; IPR001007; VWF C.  
 DR InterPro; IPR001846; VWF D.  
 DR Pfam; PF00007; Cys\_knot; 1.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; vwc; 1.  
 DR Pfam; PF00094; vwd; 4.  
 DR SMART; SM00214; VWC; 2.  
 DR SMART; SM00216; VMD; 4.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01208; VWFC\_1; 2.  
 DR PROSITE; PS0184; VWFC\_2; 2.  
 DR Glycoprotein; Repeat; Signal.  
 KW SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 5179 MUCIN 2.  
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
 FT REPEAT 1401 1416 1.  
 FT REPEAT 1417 1432 2.

FT REPEAT 1433 1448  
 FT REPEAT 1449 1464  
 FT REPEAT 1471 1478  
 FT REPEAT 1472 1478  
 FT REPEAT 1479 1494  
 FT REPEAT 1495 1517  
 FT REPEAT 1518 1533  
 FT REPEAT 1534 1556  
 FT REPEAT 1557 1572  
 FT REPEAT 1573 1596  
 FT REPEAT 1597 1612  
 FT REPEAT 1613 1635  
 FT REPEAT 1636 1651  
 FT REPEAT 1652 1675  
 FT REPEAT 1676 1683  
 FT REPEAT 1684 1699  
 FT REPEAT 1700 1715  
 FT REPEAT 1716 1731  
 FT REPEAT 1732 1747  
 FT DOMAIN 4815 4886  
 FT DOMAIN 4924 4991  
 FT DOMAIN 5075 5160  
 FT DISULFID 5075 5122  
 FT DISULFID 5089 5136  
 FT DISULFID 5098 5152  
 FT DISULFID 5102 5154  
 FT DISULFID ? 5159  
 FT CARBOHYD 153 163  
 FT CARBOHYD 423 423  
 FT CARBOHYD 670 670  
 FT CARBOHYD 770 770  
 FT CARBOHYD 894 894  
 FT CARBOHYD 1139 1139  
 FT CARBOHYD 1154 1154  
 FT CARBOHYD 1215 1215  
 FT CARBOHYD 1230 1230  
 FT CARBOHYD 1246 1246  
 FT CARBOHYD 1787 1787  
 FT CARBOHYD 1820 1820  
 FT CARBOHYD 4339 4339  
 FT CARBOHYD 4351 4351  
 FT CARBOHYD 4362 4362  
 FT CARBOHYD 4373 4373  
 FT CARBOHYD 4422 4422  
 FT CARBOHYD 4438 4438  
 FT CARBOHYD 4502 4502  
 FT CARBOHYD 4616 4616  
 FT CARBOHYD 4627 4627  
 FT CARBOHYD 4752 4752  
 FT CARBOHYD 4787 4787  
 FT CARBOHYD 4881 4881  
 FT CARBOHYD 4888 4888  
 FT CARBOHYD 4955 4955  
 FT CARBOHYD 4970 4970  
 FT CARBOHYD 5019 5019  
 FT CARBOHYD 5038 5038  
 FT CARBOHYD 5069 5069  
 FT CONFLICT 1351 1351  
 FT CONFLICT 1412 1412  
 FT CONFLICT 1449 1449  
 FT CONFLICT 1504 1504  
 FT CONFLICT 4192 4192  
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;  
 Alignment Scores:  
 Pred. No.: 2.26e-06 Length: 5179  
 Score: 253.50 Matches: 160  
 Percent Similarity: 33.29% Conservative: 76  
 Best Local Similarity: 22.57% Mismatches: 267  
 Query Match: 6.25% Indels: 206  
 DB: 1 Gaps: 34  
 US-09-914-958b-35 (1-2177) x MUC2\_HUMAN (1-5179)



YHL1\_EBV  
ID YHL1\_EBV STANDARD; PRT; 660 AA.  
AC P03181;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Hypothetical BHLFI protein.  
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10377;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84270667; PubMed=6087149;  
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
RA Tuffnell P.S., Barrell B.G.;  
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
RL Nature 310:207-211(1984).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; V01555; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A03742; Q0B63.  
KW Hypothetical protein; Early protein; Repeat.  
FT DOMAIN 149 648 4 X 125 AA TANDDEM REPEATS.  
FT REPEAT 149 273 1.  
FT REPEAT 274 398 2.  
FT REPEAT 399 523 3.  
FT REPEAT 524 648 4.  
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;  
Alignment Scores:  
Pred. No.: 5.55e-05 Length: 660  
Score: 223.50 Matches: 162  
Percent Similarity: 27.30% Conservative: 28  
Best Local Similarity: 23.28% Mismatches: 218  
Query Match: 5.41% Indels: 288  
DB: 1 Gaps: 34  
US-09-914-958b-35 (1-2177) x YHL1\_EBV (1-660)  
QY 1965 TCTTCAACAGAGGCTCCACCGCCAGGCTGTGCATCCT-----CGT 1918  
Db 130 SerAlaGlyLeuGlySerArgGlyProArgPro-----HisProAlaPheGlnValGln 147  
QY 1917 TGG---GCTCGTAGCCCGACCTCACTTCTGTGTGTGTCGGACCTGTAGTTCATAGG 1861  
Db 148 TrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArg 167  
QY 1860 TGTGTACATACGAGGCTGTCTCGAGGGGACACAGCAAGTGGCACTGCATTCATTGG 1801  
Db 167 ----- 167  
QY 1800 TGTTCATGAATGGCCATTTTCTCTGCTGGGAGAACTCTTGGCCACACTCCAGTGGGGCTG 1741  
Db 168 -----GlyHisProProGlyAla 174  
QY 1740 GGTCTGAAGTGTCTCATCCCGGGGTAGAGGATCTCCAGCACTGATTCATCTCCC 1681  
Db 175 GlyGlnArgProSerGlyProThrGlyGlyArg-----ProAlaAlaProGlyAlaPro 192  
QY 1680 CGCTGGCCAGTTCGGGCTCACCATCTTGCCATCTGGCCACGCTCACCTCCACTGCTGG 1621  
Db 193 GlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaVal---ProSerGlyAlaThr 211  
QY 1620 CTTTCATCTCTCCCGAGGCCAAAGTGTGCCAGGGGCTCCATCTCACACAGGTAGCTGAGC 1561

212 ProHisProGluArgGly---SerGlyProAlaAspProProAlaAlaAlaArgLeuPro 230  
1560 CCCCGTCGATGATCTCAGTGGGCCCCACTCTTCTTGGTGTAGACGACGACTTAGCTC 1501  
231 ProGlu-----ArgGlnGluProArgLeu 238  
1500 CCTCGCAAAGGCCCCCAACCGGTGCGTGGCACCACCTCGACGACGATTTGTTGTAAGC 1441  
239 Pro-----GlnAspLeuAlaAla-----Ala 245  
1440 CCTGATTGCCCGGAAGACGACAGCGGTGAGCCATGACTCTCTCATGGACAAGATGA 1381  
246 GlnArgCysPro-----AlaGlyProProProThrArgSerGly----- 258  
1380 GGTCCAGCATCCCGTCTCGTCCGAGTCCGTCCACACACCCCTGTGCCCC----- 1330  
259 -----AlaAlaAlaGlnArgThrHisArgArgPro---ProGlyCysProArgSerAla 275  
1329 ---GGCCCTCAGGCTCCCAAGCGGTCCGCGGATTGAGCTCTCTCGATAGGGGTCTCCGT 1273  
276 ArgAsnProGlyCysProArgThr----- 283  
1272 GCTCTCTACGGAAGACGCGGAAGAGCGGTTGGCTGAGAGCTGCGTAGGCAATTTGT 1213  
284 -----TrpArgArgArgSerGlyAlaGln----- 291  
1212 TGAAGAAGATCTCCAGCTCTCGTCTCATTTGTCAAAGTCGCGGTGATGACCTCGGACAG 1153  
292 ---ArgGlyHisProProProGlyAlaGlyGlnArg-----ProSerGlyPro 306  
1152 ---GGGAGG-----GCATGGAGAACTTGGGTGAGGCGATGTCGCGA 1114  
307 ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 326  
1113 AGCGGA-----CCTTCCCATGGTCTCATTTGTCAGATAGAGCGGTGGGGCCAT 1063  
327 GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAla 346  
1062 TCC----- 1060  
347 AspProProAlaAlaAlaArgLeuProGluArgGlnGluProArgLeuProGlnAsp 366  
1059 ---AGTTGCCATAGACGATGTCACCTTTGCCATCACGTTGAAGTCAGCCAGGCGCAC 1003  
367 LeuAlaAlaAlaGlnArgCysPro-----AlaGlyProProProThr 380  
1002 CTCGCCCATGTGCTGGGGGTGCTCCACACCGACTGCGCGGCGAGCGTCCACAAAGTGC 943  
381 ArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAla 400  
942 CATCGCCCGGTGTGGAAAGGAAGTAGGCCCATTTCTCATTTGTCGACAGATATCCG 883  
401 ArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisPro 420  
882 AGGCACTGTGCTGAGGATGGGGCCACACGCTGACGCTCGGCCCTGTATATTGCTGA 823  
421 Pro----- 421  
822 CCCGACCTCAGCAGCCACATCTCTGAGGCCAGAGATGCCCGGAGAGGTCTACTGGCCT 763  
422 -----ProGlyAlaGlyGlnArgPro 428  
762 CAGGGTCCATTTCAATGAGGGCATCAGGGCCACATTACCGTAGGGGTAAATGGCAATGT 703  
429 SerGlyPro----- 431  
702 AGATAGATGATGCTCCAGAGCCCTTTCTGTCCACAGGCCACAGAGCGTCCCGGCAAGA 643  
432 -----ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaPro 448  
642 GGTGGCCACACACCGGCCACGTTGACCTCATCTCGCTCAGGATGTCTTCCCGACGGTTAT 583

449	Gly	-----ProGlyGlyGlyAlaAlaValProSerGly	459
582	Qy	TGCGGAACCTGAAACAACCTGTGGGTGACGTGGCCACCCCGAGAAAGGCATTATTGGTGT	523
460	Db	-----AlaThrProHisProGluArg	466
522	Qy	TGAGGAAGTAGATCTTCCTCCGGCCGTCCTCCGTCGATGTCGGCAGGCTGTGACCCCAATGG	463
467	Db	-----GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArg	483
462	Qy	CGTTTCCCTGCGCGGTCCC-----CGAGCGGTAGTAGGGTGCAGCTGCCTCATCGA	412
484	Db	GlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProPro	503
411	Qy	CCGCGATGTTCAACGACCGCTTCTGGGCGCGGTCTACTTCAGAACCCAGTTGGGTCCAT	352
504	Db	ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGly	520
351	Qy	TGTACCCGCGCACGACGATCTCAAAAGTCCCCATCATGGTCCACATCAGTAATGCCACAC	292
521	Db	Cys-----ProArgSerAlaArgAsnPro	528
291	Qy	CATAGTTGAGCTGGGTGGGATTACTGTCTATAGTCAGGAGGCAGAACTGAGTTGGTGACTG	232
528	Db	-----	528
231	Qy	CAGTGAACATGGTTTCAGCCCGCTGGGACCCCTCAGTGATGGCGAGAACCCAGAGCAGCA	172
529	Db	-----GlyCysProArgThrThrTipArgArgArgSerGlyAla	540
171	Qy	GCAGGAACGCTAACATCTCGGATGCCGGGTACGCGCTCGAGGCCATCTCCCGCTCT	112
541	Db	-----GlnArgGlyHisPro-ProProGlyAla	549
111	Qy	CGGCGCCCGCTAGGGCGGTGGGAAGCGGCGCTCGCTGCGCGCTCTGCGCCCGCGCGC	52
549	Db	ArgGlnArg-ProSerGlyProThrGlyGlyArg-----ProAlaAlaProGlyAla	567
51	Qy	CG-----CGCCCTGTGCTCCAGCCCGGTCCCGGGCTGGCC	13
567	Db	roGlyThrProAla-AlaProGlyProGlyGlyAlaAla	580

PROFIT E C

RESOLX	HSVEB	STANDARD;	PRT; 797 AA.
ID	P28968;		
AC	01-DEC-1992 (Rel. 24, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DT	01-DEC-1992 (Rel. 24, Last annotation update)		
DE	Glycoprotein X precursor.		
GN	71.		
OS	Equine herpesvirus type 1 (strain Ab4p) (EHV-1).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Alphaherpesvirinae; Varicellovirus.		
OX	NCBI_taxid=31520;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92295566; PubMed=1318606;		
RA	Telford E.A.R., Watson M.S., McBride K., Davison A.J.;		
RT	"The DNA sequence of equine herpesvirus-1.";		
RL	Virology 189:304-316(1992).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; M86664; AB02506.1; -.		
DR	PIR; H36802; VGBEX1.		
DR	InterPro; IPR007110; IG-like.		

```

CC 1243 ACCGCTCTTCGCGTCATCCGTA---GAGACACGGAGACCCCTCATCGAGGACTCA 1299
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 264 ThrSerGlySerThrThrThrThrGlyAlaSerThrThrProSerAlaSerThrAla 283
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1300 ATCCCGGCGACGCTGGAGCCTGAGCGCGGGGCAAGGGGTTGTGGTACCGACTTCG 1359
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 284 ThrSerAlaThrProThrSer-----ThrSer 292
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1360 ACGGACGGGCTGGGACCTCATCTTCCATGGAGAGTCCATGGCTCAGCCGCTGT 1419
CC ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 293 ThrSerAlaAlaThrThrSer-----ThrProThr 303
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1420 CGCTCTTCGCGGCAATCAGGCTTCAACAACATGGCTGGAGTGGTGGCAGCACCC 1479
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 304 ProThrSerAlaAlaThrSerAlaGluSerThr-----GluAlaPro 318
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1480 GGTGGGCGCTTGGCCAGGGAGCTAAGTGTCTACACCAAGAGTGGGGCC 1539
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 319 -----ThrSerThr-----Pro 322
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1540 ACCTGAGGATCAGCGGGGCTCAGGCTACCTGTGTGAGATGGAGCCGTGGCACACT 1599
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 323 ThrThrAspThrThrThrProSerGluAlaThrThrAlaThrThrSerProGluSerThr 342
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1600 TTGGCTGGGGAAGATGAAGCCAGCAGTGTGGAGTGACGTGGCCAGATGGCAAGATGG 1659
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 343 ThrValSerAla-----346
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1660 TGAGCCGGAACTGGCCAGCGGGAGATGAATCAGTGTGGAGATCTCTACCCCGGG 1719
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 347 -----SerThrThrSerAla 351
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1720 ATGAGGACACACTTCAGGACCCAGCCCACTGAGTGTGGCAAGATCTCCAGCAGG 1779
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 352 ThrThrThrAlaPheThrThrGluSerHisThrSer-----363
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1780 AAATGCGCATTCATGGACACCAATGATGATCCAGTCCAGTTCGCTCGCTCGAG 1839
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 364 -----ProAspSerSerThrGlySerThrThrAlaGluPro 376
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1840 ACAAGCCCGTATGTCTACACCTATGGAAGTACAGGTGCGGACCAACAAGAGTGCA 1899
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 377 SerSerThrThrLeuThrProSerThrAlaThrProSer-----ThrAspGlnPhe 394
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 1900 GTCGGGCTACGAGCCCAACG 1920
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CC 395 ThrGlySerSerAlaSerThr 401
CC |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
ID YS89 CAEEL STANDARD; PRT; 3178 AA.
AC Q09624; Q09625; Q096D4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson-Spratt J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Contains 1 GPS domain.

```

```

-!- SIMILARITY: Contains 1 PLAT domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48544; CAB70192.1; -
CC EMBL; Z48582; CAB70192.1; JOINED.
CC EMBL; Z48582; CAB7020.1; -
CC EMBL; Z48544; CAB7020.1; JOINED.
CC WormPep; ZK945.9; CE25697.
CC InterPro; IPR002111; Cat_channel_TrpL.
CC InterPro; IPR005821; Ion_trans.
CC InterPro; IPR001024; Lipoxigenase_LH2.
CC InterPro; IPR003915; PKD_2.
CC InterPro; IPR002023; PKD_cys_rich.
CC InterPro; IPR008976; PLAT_LH2.
CC Pfam; PF01825; GPS; 1.
CC Pfam; PF00520; Ion_trans; 1.
CC Pfam; PF01477; PLAT; 1.
CC PRINTS; PR01433; POLYCYSTIN2.
CC SMART; SM00303; GPS; 1.
CC SMART; SM00308; LH2; 1.
CC PROSITE; PS00095; PLAT; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 13 30
FT TRANSMEM 51 73
FT TRANSMEM 2139 2161
FT TRANSMEM 2348 2367
FT TRANSMEM 2390 2412
FT TRANSMEM 2451 2468
FT TRANSMEM 2483 2505
FT TRANSMEM 2567 2589
FT TRANSMEM 2836 2858
FT TRANSMEM 2939 2961
FT TRANSMEM 2976 2998
FT TRANSMEM 3038 3060
FT DOMAIN 256 1196
FT DOMAIN 1105 1241
FT DOMAIN 2071 2120
FT DOMAIN 2182 2305
FT DOMAIN 3178 AA; 34726 MW; F8239436D03666CD CRC64;
SQ SEQUENCE 3178 AA; 34726 MW; F8239436D03666CD CRC64;

Alignment Scores:
Pred. No.: 0.00105 Length: 3178
Score: 203.50 Matches: 151
Percent Similarity: 36.97% Conservative: 83
Best Local Similarity: 23.85% Mismatches: 210
Query Match: 5.02% Indels: 189
DB: 1 Gaps: 28

US-09-914-958B-35 (1-2177) x YS89_CAEEL (1-3178)
QY 190 CCATCACTGAGG-----GGTCCAGCGGGCTGAACCCATGT-----TCACCTGCAG 234
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 253 ProThrLeuArgMetLysArgAspAlaGlyAspAsnThrCysAspThrThrIleGlu 272
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 235 TCACCAACTCAGTTCTGCTCTGCTCTGCTATGACAGTAATCCACCCAGCTCAACTATGG 294
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 SerThrSer-----ThrSerThrThrThrProThrThrThrThrValThr 287
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 295 TGGCAGTTACTGATGTGACCATGATGGGAGCTTTGAGATCGTCTGGCGGGGTACAATG 354
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 288 -----SerThrValThrSerThr-----ThrThr 295
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 GACCAACCTGGTTCTGAGTATGACCGGCCCAAGCGGCTGGTGAACATCGCGGTG 414
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 296 ValProThrSerThrSerThrValThr-----ThrAlaMetSer 308
Db |||: ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; V01555; CAA24854.1; -.  
 DR PIR; B43042; Q0BE21.  
 DR PIR; B43042; Q0BE22.  
 DR InterPro; IPR007796; Herpes BLF1.  
 DR Pfam; PF05109; Herpes BLF1.  
 KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.  
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 435 435 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 589 589 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 610 610 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 656 656 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 683 683 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 746 746 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 815 815 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 858 858 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 888 888 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARSPLIC 502 698 Missing (in isoform GP220).  
 FT /FTid=VSP\_002070.  
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC52C9 CRC64;

## Alignment Scores:

Pred. No.: 0.00159 Length: 907  
 Score: 197.50 Matches: 126  
 Percent Similarity: 36.47% Conservative: 56  
 Best Local Similarity: 25.25% Mismatches: 216  
 Query Match: 4.87% Indels: 102  
 DB: 1 Gaps: 23

US-09-914-958b-35 (1-2177) x VGP3\_EBV (1-907)

QY 600 ATCTGAGCGATGAGTCAAGTGGCGCCGTGGTGGCCACACCTCTTTGGCGGAC----- 654  
 Db 304 ValPheSerAspGluileProAlaSerGlnAspMet-ProThrAsnThrThrAspIleTh 323  
 QY 655 -----GCTCTGGCTGTGTGACAGAGAGGCTCTGGACGCTACTCT 698  
 Db 323 rTyValGlyAspAsnAlaThrTySerValProMetValThrSerGluAspAla----- 341  
 QY 699 ATCTACATTGCAATTACGCTACGGTAAATGTGGGCGCTGATGCCCTCA----- 747

Db 342 -AsnSerProAsnValThrValThrAlaPheTrpAla-----TrpProAsnAsnThrGluTh 360  
 QY 748 -----TTCAATGAGCCCTGAGCCAGTGCCTCTCCGGGGCATTTGGCGCTC 797  
 Db 360 rAspPheLysCysLysTrpThrLeuThrSerGlyThrProSerGlyCys----- 376  
 QY 798 AGAGATGTGGCTGCTGAGGCTGGGTGAGCAATAATACAGGGGGCGGAGCGTCA----- 852  
 Db 377 -----GluAsnIleSerGlyAlaPheAlaSerAsnAr 387  
 QY 853 -----GCGTGGGCCCCATCTCTCAGCAGCAGTGCCTCG 884  
 Db 387 gThrPheAspIleThrValSerGlyLeuGlyThrAlaProLysThrLeuIleThrAr 407  
 QY 885 GATATCTTCTGCGACAAATGAGAAATGGCCCTTCTTCCACAAACCGGGCGCATGGC 944  
 Db 407 gThrAlaThrAsn-----AlaThrThrThrHisLys 418  
 QY 945 ACCTTTGTGGACGCTGCGGCGCAGTGTGTGTGGACGCCCAACC-----AGCATGGG 998  
 Db 418 sValIlePheSerLysAlaProGluSerThrThrSerProThrLeuAsnThrThrGl 438  
 QY 999 CGAGGTGTGCGCCCTGGCTGACT-----TCACCGTGTGATGGCAAGTG 1040  
 Db 438 yPheAlaAspProAsnThrThrThrGlyLeuProSerSerThrHisValProThrAsnLe 458  
 QY 1041 GACATCGTCTATGGCAACTGGA-----ATGGCCCCCAGCTCTATCTGCAATGAGCACC 1097  
 Db 458 uThrAlaProAlaSerThrGlyProThrValSerThrAlaAspValThrSerProThrPr 478  
 QY 1098 CATGGGAAGTCTCGCTTCGGGACATCGCTCACCACCAAGTTCTCCATGCGCT-----CCCT 1154  
 Db 478 oAlaGly---ThrThrSerGlyAlaSerProValThrProSerProSerProThrAspAs 497  
 QY 1155 GTCCGACGCTATACCGCGGACTTGTGACATGACGAGGAGTGGAGATCTTCTTCAAC 1214  
 Db 497 nGlyThrGluSerLysAlaProAspMetThr-----SerSe 509  
 QY 1215 ACATTTGCGTACCGCAGCTCTCAGCCACCGCTCTTCGCGCTCATCCGTAGAGCAC 1274  
 Db 509 rThrSerProValThrThrProThrProAsnAlaThrSerProThrProAlaValThrTh 529  
 QY 1275 GGAGACCCCTCATCGAGGAGCTCAATCCCGCG-----ACGCTTGGAGCCT 1322  
 Db 529 rProThrProAsnAlaThrSerProThrProAlaValThrThrProThrPro-----AsnAl 548  
 QY 1323 GAGGGCGGGGACAGGGGTGTGTGACCGACTTCGACGAGCAGGAGGATGCTGGACCTC 1382  
 Db 548 aThrSerProThrLeuGlyLysThrSerProThrSerAlaValThr-----ThrPr 565  
 QY 1383 ATCTTGTCCCATGGAGATCCATGGCTCAGCGCTGTCGCTTTCGGGGCAATCAGGGC 1442  
 Db 565 oThrProAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAlaValThrTh 585  
 QY 1443 TTCACAAACAACCTGGCTCGAGTGGTGCACGACCCCGTTGGGGCTTTG-----CCAGG 1499  
 Db 585 rProThrProAsnAlaThr-----SerProThrLeuGlyLysThrSerProTh 601  
 QY 1500 GGAGCTTAAGTGTGTCTTACACCAAGAGTGGGCGCCACCTGAGGATCATCGAGGG 1559  
 Db 601 rSerAlaValThrThrProThrProAsnAlaThrGlyProThrValGly---GluThrSe 620  
 QY 1560 GGCTCAGGCTACCTGTGTGAGATGGAGCGCTGACACACTTTGGCTTGGGGAAGATGAA 1619  
 Db 620 rProGlnAlaAsnAlaThrAsn-----HisThrLeuGlyGlyThrSerProTh 636  
 QY 1620 GCCACGAGTGTG-----GAGTGAAGTGGCCAGATGGCAA 1654  
 Db 636 rProValValThrSerGlnProLysAsnAlaThrSerAlaValThrThrGlyGlnHisAs 656  
 QY 1655 rATGCTGAGCGCGAACGTGGCCACGCGGGAGATG-----AACTCAGTGTGGAGATCCT 1708  
 Db 1708

```

656 nileThrSerSerThrSerSerMetSerLeuArgProSerSerAsnProGluThrLe 676
1709 CTACCCCGGATGAGGACACACTTCAGGACCCAGCCCATGAGTGTCGCCAAGGATT 1768
676 uSerProSerThrSerAspAsnSerThrSerHisMetProLeuLeuThrSerAlaHisPr 696
1769 CTCCAGCAGGAAATGCGCCATTCATGAGCAGCACAATGAATGCATCCATGCCATTCGT 1828
696 oThrGlyGluAsn-IleThrGlnValThrProAlaSerIleSerThrHisHisVals 716
1829 GTGCCCTCGAGACAGAGCCGCTATGTCTCAACACCTATGGAAGCTACAGTCCCGGA--- 1884
716 erThrSerSerProAlaProArgProGlyThrThrSerGlnAlaSerGlyProGlyAsnS 736
1885 --CCAAACAAGAGTGCAGTCGGGCTACGAGCCCAAGAGGATGGCACAGCT 1935
736 erSerThrSerThrLysProGlyGluValAsnValThrLysGlyThrProPro 753

RESULT 8
P121 HUMAN
ID P121 HUMAN STANDARD; PRT; 1229 AA.
AC Q9F2N3; O75115; Q9Y4S7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE of 121 kDa) (P145).
GN POM121 OR NUP121 OR KIAA0618.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cordes M., Bauer C., Holmes A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 243-1229 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [3]
RP SEQUENCE OF 1130-1229 FROM N.A.
RC TISSUE=Uterus;
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Essential component of the nuclear pore complex. The
CC repeat-containing domain may be involved in anchoring components
CC of the pore complex to the pore membrane (By similarity).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE (BY SIMILARITY).
CC -!- DOMAIN: Contains F-X-F-G repeats.
CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AC006014; AAD28064.1; -
CC EMBL; AB014518; BAA31593.1; -
CC EMBL; AL080109; CAB45713.1; -
CC PIR; T12551; T12551.

```

```

DR KW Genew; HGNC:19702; POM121.
FT DOMAIN 1 40 Nuclear protein; Transport; Transmembrane; Repeat.
FT TRANSMEM 41 61 CISTERNAL SIDE (POTENTIAL).
FT DOMAIN 62 1229 POTENTIAL.
FT DOMAIN 4 10 PORE SIDE (POTENTIAL).
FT DOMAIN 51 56 POLY-ALA.
FT DOMAIN 294 299 POLY-LYS.
FT DOMAIN 441 444 POLY-SER.
FT DOMAIN 499 502 POLY-PRO.
FT DOMAIN 733 736 POLY-SER.
FT DOMAIN 819 826 POLY-SER.
FT DOMAIN 869 875 POLY-SER.
FT DOMAIN 1061 1067 POLY-THR.
FT CONFLICT 297 297 K -> E (IN REF. 2).
FT CONFLICT 336 336 S -> N (IN REF. 2).
FT CONFLICT 379 379 T -> A (IN REF. 2).
FT CONFLICT 451 451 K -> R (IN REF. 2).
FT CONFLICT 545 545 T -> I (IN REF. 2).
FT CONFLICT 559 559 T -> P (IN REF. 2).
FT CONFLICT 702 702 S -> P (IN REF. 2).
FT CONFLICT 881 881 P -> H (IN REF. 2).
FT CONFLICT 965 965 T -> A (IN REF. 2).
FT CONFLICT 982 984 PAT -> A (IN REF. 2).
FT CONFLICT 991 996 ASTIKI -> PSMIKV (IN REF. 2).
FT CONFLICT 1000 1000 H -> Y (IN REF. 2).
FT CONFLICT 1006 1008 OPT -> HPI (IN REF. 2).
FT CONFLICT 1044 1044 MISSING (IN REF. 2).
FT CONFLICT 1102 1102 A -> T (IN REF. 2).
FT CONFLICT 1124 1124 T -> A (IN REF. 2).
FT CONFLICT 1146 1146 S -> G (IN REF. 1).
FT CONFLICT 1165 1165 L -> Q (IN REF. 1).
FT CONFLICT 1195 1195 A -> G (IN REF. 1).
FT CONFLICT 1199 1229 SAAPSFSGASGKTPGARQLQARQROHTRKK -> NTFARQ
OEHSPPKGNLNSKRKLLPAVRAGGPPRRGQASFPTRKE
(IN REF. 2).
FT CONFLICT 1202 1202 P -> L (IN REF. 3).
FT CONFLICT 1229 1229 SQ SEQUENCE 1229 AA; 125087 MW; 515655D1285898B CRC64;

Alignment Scores:
Pred. No.: 0.00184
Score: 197.00
Percent Similarity: 33.10%
Best Local Similarity: 22.62%
Query Match: 4.86%
DB: 1

US-09-914-958B-35 (1-2177) x P121_HUMAN (1-1229)
QY 47 CGCGCGCGCGCGCGAGAGCGCGCGCTTCCACGCGCCCTAGCGCGGG 106
Db 464 ArgLysLysGlnAsnSerAsnSerGlnSerThrProGlySerSerGlyGlnArgLys--- 482
QY 107 GCGCGAGAGCGGAGGATGGCTCCGAGCGCTGACCCCGGCGATGTCCAGGANGTACCGTT 166
Db 483 -----ArgLysValGln-LeuLeuPro-- 489
QY 167 CTGTGCTGCTCTGTGTTTCTGCCCATCACTAGGAGGGTCCCGAGGGCTGAACCCATGTT 226
Db 490 -----SerArgGlyGlnLeu-- 496
QY 227 CACTGCAGTCACCACTCACTTCTGCTCTGACTATGACAGTAATCCACCCAGCTCAA 286
Db 497 -----ThrLeuProPro-----ProGlnLeuG1 505
QY 287 CTATGTGTGGCGATTACTGTGTGACCATGATGGGACTTTGAGATCGTCTGGCGGG 346
Db 505 yTyrSerIleThrAlaGluAspLeuAspLeuGlu----- 516
QY 347 GTACATGGACCCCACTCGTTCTGAGTATGACCGGGGCCAGACGGGCTGTGACAT 406
Db 517 -----LysLysAlaSerLeuGlnTrpPheAsnG1 526

```





```

QY 193 TCACTGAGGGT-----CCACGGGCGCTGAAC----- 219
Db 192 LysGlyHisGlyGlyIleHisCysPheHisArgLysProLysGlyHisSerHisGluGlu 211
QY 220 -----CCATGTCTCACTCAGTCAACCACTCAGTTCTCCCTCCTCAGTATGACAGTAATC 273
Db 212 HisThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 231
QY 274 CCACCCAGCTCAACTATGTTGGCGAGTTACTGATGTGACCAATGATGGGAGCTTTGAGA 333
Db 232 ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 251
QY 334 TCGTCTGGGGGTACAATGACCAACCTGGTCTCGAATATACCGGGCCCAAGAC 393
Db 252 Thr-----ProThrThr 255
QY 394 GCGTGTGAACATCGCGTGCATGAGCGCAGCTCACTTACCGCTCGCGGACCGGC 453
Db 256 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 275
QY 454 AGGGAAAGCCATGCGGTTCACAGCTCGACATCGAGGGGACGGCCGGGAGGAGATCT 513
Db 276 LysAlaThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 295
QY 514 ACTTCTCAACACCAATATGCTTCTCGGGGTGCCACGTTACACCGACA----- 564
Db 296 AlaThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 315
QY 565 AGTTCTCAAGTTCGCGCAATACCGGTGGGAAGACATCTGAGCGATGAGGTCAACGTGG 624
Db 316 AspCysGlyTyrSerGlyIleThrGluSerGlnCys-----ArgThrLys 330
QY 625 CCCGTGTGTGCCAGCTCTTTGCGGACGCTCTGTGSCCTGTGTGACAGAAAGGCT 684
Db 331 GlyCysCysPheAspSerSerileProGlnThrLysTyrCysPheTyrThrLeu----- 348
QY 685 CTGGAGCTACTCTACTATCTACATTCGAATTAGCGCTACGTTAATGTGGCCCTGTATGCC 744
Db 349 -----SerGlnValAlaAspCysLysValGluProSer-----Gln 360
QY 745 TCATTGMAATGACCTCAGGCGCAGTCACTCTCCCGGGGCATCTTGGCGCTCAGAGATG 804
Db 361 ArgValAspCysGlyPheArgGlyIleThrAlaAspGlnCysArgGlnLysAsnCysCys 380
QY 805 TGGCTGTGAGGCTGGGTTCAGCAATATACAGGGGCGCAGCGGTACGCTGGGCCCCA 864
Db 381 Phe-----AspSerSerileSerGlyThrLys-----TrpCysPhe 392
QY 865 TCCTCAGCAGCAGTCCCTCGGATATCTCTGCGACAATGAGAAATGGCCCTAATCTCTTT 924
Db 393 TyrSerThrSerGlnValAlaAlaThrLysThrThrThrThrThrThrThrThrThrThr 412
QY 925 TCCACAAACCGGGCGCATGGACCTTTGTGGACGCTGCGGCGCAGTGTGTGTGGACGACC 984
Db 413 ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 432
QY 985 CCACACAGATGGGAGAGTGTGCGCTGTGCTGACTTCAACCGTGTGATGGCAAGTGACA 1044
Db 433 ProThrThrThr-----ThrThrThrThrThrThrThrThrThrThrThrThrThr 446
QY 1045 TCGTCTATGCCATGGAAATGGCCCGCCCGCCCTCTATCTGCAAAATGAGACCCATGGGA 1104
Db 447 ThrThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 461
QY 1105 AGGTCCGCTTCCGGGACATCGCTCAACCCCAAGTTCTCCATGCGCTCCCTGTCGCGACG 1164
Db 462 -----AlaThrThrThrPro-----ThrThrThrThrThrThrThrThrThrThr 470
QY 1165 TCATCACCGCGACTTGTGACATGACACGAGCTGGAGATCTTCTTCAACAACTGCGCT 1224
Db 471 ThrThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 489

```

```

QY 1225 ACCGAGCTCCTCAGCAACCGCCTCTCCCGCTCATCCGTAGAGACGAGACCCGCC 1284
Db 490 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 509
QY 1285 TCATCGAGAGCTCAATCCCGCGCAGCCTTGG----- 1317
Db 510 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 529
QY 1318 ---AGCTGAGGGC-----GGGCGACAGGGGTGTGGTACCGACTTCGACGAGAGCG 1368
Db 530 ProSerLysArgAlaAspCysGlyTyrProGlyIle-----ThrGluSer 544
QY 1369 GCATCTCGACTCATCTTGTCTCCATGGAGAT----- 1401
Db 545 GlnCysArgSerLysGlyCysPheAspSerSerileProGlnThrLysTrpCysPhe 564
QY 1402 -----CCATGGCTCAGCGCTGT-----CCGTCTTTC----- 1428
Db 565 TyrSerLeuProGlnValAlaAspCysLysValAlaProSerSerArgValAspCysGly 584
QY 1429 ---GGGCAATCAGGGCT-----TCAACAACA 1452
Db 585 PheGlyGlyIleThrAlaAspGlnCysArgGlnArgAsnCysCysPheAspSerSerile 604
QY 1453 ACTGCTCGAGTGTGTGC-----CACGCAACCGGTTTGGGGCT 1491
Db 605 SerGlyThrLysTrpCysPheTyrSerThrSerGlnGlyAsnAlaMetCysSerGlyPro 624
QY 1492 TTGCGAGGGGAGCTAAGTGTGTCTTACACCAAGAGAGTGGGCGCCCTGAGATCA 1551
Db 625 ---ProThrLysArgAspCysGlyTyrPro-----GlyIle 636
QY 1552 TCGACGGGGCTCAGCTACCTGTGTGAGA-----TGGAGC----- 1587
Db 637 SerSer-----SerValCysIleAsnArgGlyCysCysTrpAspAsnSerValMet 653
QY 1588 -----CCGTGG 1593
Db 654 AsnValProTrp 657

```

RESULT 10  
 PYGO DROME  
 ID PYGO DROME STANDARD; PRT; 815 AA.  
 AC Q9V9W8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Pygopus protein (Gammy legs protein).  
 GN PYGO OR GAM OR CG11518.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid:7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21952490; PubMed=11955446;  
 RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
 RA Murone M., Zuehlig S., Basler K.;  
 RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
 RT pygopus to the nuclear beta-catenin-TCF complex.";  
 RL Cell 109:47-60(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Head;  
 RX MEDLINE=22010046; PubMed=12015286;  
 RA Parker D.S., Jemison J., Cadigan K.M.;  
 RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling  
 RT in Drosophila.";  
 RL Development 129:2565-2576(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;







DR Pfam; PF03093; Nucleoporin\_FG; 8.  
 KW Nuclear protein; Transport; Transmembrane; Repeat.  
 FT DOMAIN 1 56 CISTERNAL SIDE (POTENTIAL).  
 FT TRANSMEM 57 77 POTENTIAL.  
 FT DOMAIN 76 1199 PORE SIDE (POTENTIAL).  
 FT DOMAIN 4 8 POLY-ALA.  
 FT DOMAIN 53 58 POLY-ALA.  
 FT DOMAIN 438 441 POLY-SER.  
 FT DOMAIN 497 500 POLY-PRO.  
 FT DOMAIN 684 687 POLY-SER.  
 FT DOMAIN 1023 1026 POLY-ALA.  
 FT DOMAIN 1033 1038 POLY-SER.  
 FT DOMAIN 1058 1063 POLY-GLY.  
 SQ SEQUENCE 1199 AA; 120784 MW; 6DC4451B91D5B907 CRC64;

Alignment Scores:  
 Pred. No.: 0.00364 Length: 1199  
 Score: 191.50 Matches: 190  
 Percent Similarity: 32.86% Conserv: 63  
 Best Local Similarity: 24.68% Mismatches: 307  
 Query Match: 4.72% Indels: 210  
 DB: 1 Gaps: 35

US-09-914-958b-35 (1-2177) x P121\_RAT (1-1199)

QY 20 CCGGGACCGGGCTGGGAGCAAGCAGCGCGGGCGCGCGG-----CAGAGG 67  
 Db 471 ProThrProGlySerSerGlyGlnArgLysArgLysIleGlnLeuLeuProSerArgArg 490  
 QY 68 CGCAGCGAGCGCGCTCCCGCCCGCTAGCGGGCGGGCGGAGAGCGGAGGATGCG 127  
 Db 491 -GlyAspGlnLeuThrLeuProPro-----ProGluLeuGlyTyrSerII 506  
 QY 128 TCGAGGGCTGACCCGGGATGTCAGAGATGTTACCGTTCCTGCTGCTCTGGTT-- 185  
 Db 506 eThraGluAspLeuMetGluArgAlaSer-----LeuGlnTrpPheAs 523  
 QY 186 -----CTGCCCATCAGTGGGGTCCCGCG 211  
 Db 523 nLysValLeuGluAspLysThrAspAspAlaSerThrProAlaThrAspThrSerProAl 543  
 QY 212 GGCTGAACCCATGTCACGTGACGTACCACTCAGTTCTGCTCT----- 254  
 Db 543 aThrSerProPheThrLeuThrLeuProThrValGlyProAlaAlaSerProAlaSe 563  
 QY 255 ----CTGACTAGCAGTAATCCC-----ACCAGCTCAACTATGTTGT 295  
 Db 563 rLeuProAlaProSerSerAsnProLeuLeuGluSerLeuLysMetGlnGluSerPr 583  
 QY 296 GCGAGTTACTGATGGACCATGATGGGACTTTGAGATCGTCTGGCGGGTACATGG 355  
 Db 583 oAlaProSer-----SerSerGluProProGluAlaAlaThrValAlaAlaProSerPr 601  
 QY 356 ACCCAAC-----CTGTTCTGAAGTATGACCGGGCGCGCCAGAGCG 394  
 Db 601 oProLysThrProSerLeuLeuAlaProLeuValSer-ProLeuThrGlyPro----- 618  
 QY 395 GCTGGTGAACATCGCGGTGATGAGCGAGCTCAACCTACTAGCGC-----T 442  
 Db 619 -----LeuAlaSerThrSerSerAspSerLysProThrThrPheLeuGlyLeuA 636  
 QY 443 GCGGACCGGCGGAGGAGCCATTCGGG-----TCAGACCTG 481  
 Db 636 lAsrAlaSerSerAlaThrProLeuThrAspThrLysAlaProGlyValSerGlnAlaG 656  
 QY 482 CGACATCGAGCGGCGCGGAGGAGATCTACTCTCAACACCAATAATGCTTCTC 541  
 Db 656 lNLeuCysValSerThrProAlaAlaThrAlaProSerProThrProAlaSerThrLeuP 676  
 QY 542 GGGGTGGCCAGTACCGCAAGATGTTCCGAGTTCCGCAATAACCGGTGGGAGACAT 601  
 Db 676 heGlyMetLeuSerProProAlaSerSerSerSerLeuAlaThrProGlyProAlaCysA 696

QY 602 CCGAGCGATG---AGGTCAACGTGGCCCGCTGTGGCCAGCCTCTTTGGCGGAGCGTC 658  
 Db 696 lAsrProMetPheLysProIlePheProAlaThr---ProLysSerGluSerAspAsnP 715  
 QY 659 TGTGGCCTGTGTGACAGAAAGGCTCTGGAGCTACTCTATCTACA---TTGCAATA 715  
 Db 715 rLeuProThrSerSerAlaAlaThrThrProAlaSerThrAlaLeuProThrT 735  
 QY 716 CGCCTACGGTAATG-----TGGGCC 736  
 Db 735 hrAlaThrAlaThrAlaHisThrPheLysProIlePheGluSerValGluProPheAla 755  
 QY 737 TGATGCC-----TCATTGAATGACCTCGAGCCAGTGAACCTCTCCG 781  
 Db 755 lMetProLeuSerProPheSerLeuLysGlnThrThrAlaProAlaThrThrAla 775  
 QY 782 GGGCATCTCGCGCTCAGAGATGGCTGTGAGGCTGGGTGAGCAAT-----A 832  
 Db 775 lThr-----SerAlaProLeuLeuThrGlyLeuGlyThrAlaThrSerThrVal 792  
 QY 833 TACAGGGCGCGGCGCTCAGCGTGGGCCCA---TCCTCAGCAGCAGTGCCTCGGAT 889  
 Db 792 lThrGlyThrThrAlaSerAlaSerLysProValPheGlyPheGlyValThrThrAla 812  
 QY 890 CTTCGGGCAATGAGAAATGGCTAACTTCTTTTCCACACCGGGCGATGACCTT 949  
 Db 812 lAsrThrAlaSerThrIleAlaSerThrSerGlnSerIleLeuPheGlyGlyAlaPro 831  
 QY 950 TGTGGAGCTGCGCGCAGTCTGCTGTGGAGACCCACAGCATGGGCGAGGTGCGC 1009  
 Db 832 -----ProValThrAlaSerSerSerAlaP 840  
 QY 1010 CTTGGCTGACTTCAACCGTGTATGGCAAGTGGACATCGTATATGGCACTGGATGCC 1069  
 Db 840 roAlaLeuAlaSerIlePhe-----GlnPheGlyLysProLeuAlaP 854  
 QY 1070 CACCGCTCTATCTCAAAATGAGCACCCATGGGAAGTCCGCTCCGGGACATCGCTC 1129  
 Db 854 roAlaAlaSerVal-----AlaGlyThrSerPheS 864  
 QY 1130 ACCCAGTTCTCCATGCCCTCCCTGTCGCGCAGGTCTATCACCCTG----- 1176  
 Db 864 erGlnSerLeuAlaSerSerAlaGlnThrAlaAlaSerAsnSerSerGlyPheSerG 884  
 QY 1177 -----ACTTTGACAATGACCGAGCTGGAGATCTTCTCAACACATTCGCTA 1225  
 Db 884 lPheGlyGlyThrLeuThr-----ThrSerThrSerAlaProA 897  
 QY 1226 CCGAGCTCCTCAGCAACCGCTCTTCCGCTCATCCGTAGAGACAGC---GAGACCC 1282  
 Db 897 lThrThrSerGlnProThrLeuThrPheSerAsnThrValThrProThrPheAsnIleP 917  
 QY 1283 CTTCTACGAGGAGCTCAATCCCGGCGACCCCTTGGAGCTGAGGCGGGGCGACAGGG 1342  
 Db 917 rPheSerAlaSerAlaLysProAlaLeuProThrTyrProGlyAlaAsnSerGlnPro 937  
 QY 1343 TGTGTGACCGACTTCGACGAGAGCGGATGTGTGGACCTCATCTTCCCATGAGAGTC 1402  
 Db 937 hrPheGlyAlaThrAspGlyAlaThr-----LysP 947  
 QY 1403 CATGCTCAGCGCTGTCCGTCTCCGGGCAATCAGGCTTCAACACCACTGGCTGCG 1462  
 Db 947 roAlaLeuAla-----ProSerPheGly-----SerSerPheThrPheGlyAsnS 962  
 QY 1463 AGTGTGCCACGACCCGGTTGGGGCTTTGCCA----- 1497  
 Db 962 erValAlaSerAlaProSerAlaAlaProAlaProAlaAlaPheGlyGlyAlaAlaGlnP 982  
 QY 1498 -----GGGAGCTAAGTCTGCTCT-----ACACCAAGAGTGGGCG 1537  
 Db 982 roAlaPheGlyGlyLeuLysAlaSerAlaSerThrPheGlyThrProAlaSerThrGlnP 1002  
 QY 1538 CCACCTGAGGATCATCGAC----- 1557

```

Db 1002 roAlaPheGlySerThrThrSerValPheSerPheGlySerAlaThrThrSerGlyPheG 1022
QY 1558 -----GGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCCTGGGGA 1612
Db 1022 lyAlaAlaAlaThrThrGlnThrThrHisSerGlySerSerSerLeuPheGlys 1042
QY 1613 GGATGAAGCCAGCAGTGTGGAGGTGACGTGCCCGCAGTGGCGAAGATGGTGAGCCGGAACGT 1672
Db 1042 erSerThrProSerProPheThrPheGlyGlySerAlaAla----- 1055
QY 1673 GCGCAGCGGAGAGTAACTCAGTCTCGAGATCTCTACCCCGGGATGAGGACACACT 1732
Db 1056 --ProAlaGlyGly---GlyGlyPheGlyLeuSerAlaThrProGlyThr-GlySerThr 1073
QY 1733 TCAGGACCCAGCCCACTGGAGTGTGGCCAA----- 1763
Db 1074 SerGlyThrPheSerPheGly-SerGlyGlnSerGlyThrGlyThrThrThrSerPh 1093
QY 1764 -----GGATCTCCAGCAGGAGAAATGGCCATTGCAT 1795
Db 1093 eGlyGlySerLeuSerGlnAsnThrLeuGly-AlaProSerGlnSerSerProPheAlaP 1113
QY 1796 GGACACCAATGAATGCATCCAGTTCCTCCATTCGTGCTCGCCCTCGAGACAAGC---CGTATG 1852
Db 1113 heSerValGlySerThrProGluSerLysProValPheGlyThrSerThrProThrP 1133
QY 1853 TGTCAACACCTATGAAGCTACAGTGGCGGACCAACAGAGTGCAGTCCGGGTACGA 1912
Db 1133 heGlyGlnSerAlaProAlaProGlyValGlyThrThrGlySerSerLeuSerPheGlyA 1153
QY 1913 GCCCAACGAGATGGCAGAG 1932
Db 1153 laProSerThrProAlaGln 1159

RESULT 12
MUC1 MESAU
ID MUC1 MESAU STANDARD; PRT; 676 AA.
AC Q60528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor.
OS MUC1.
GN Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tracheal epithelium;
RX MEDLINE=96326118; PubMed=8703480;
RA Park H., Hyun S.W., Kim K.C.;
RT "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
cells in primary culture."
RL Am. J. Respir. Cell Mol. Biol. 15:237-244 (1996).
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CYTOSKELETON (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36918; AAB53965.1; -.
CC InterPro; IPR000082; SEA_domain.

```

```

DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
FT REPEAT. 1 25
FT SIGNAL. 1 25
FT CHAIN. 26 676
FT DOMAIN 26 582
FT TRANSMEM 583 603
FT DOMAIN 604 676
FT DOMAIN 458 573
FT CARBOHYD 291 291
FT CARBOHYD 323 323
FT CARBOHYD 350 350
FT CARBOHYD 380 380
FT CARBOHYD 400 400
FT CARBOHYD 413 413
FT CARBOHYD 435 435
FT CARBOHYD 479 479
FT CARBOHYD 496 496
FT CARBOHYD 536 536
FT CARBOHYD 536 536
SQ SEQUENCE 676 AA; 67616 MW; 95F479B6EC5C3884 CRC64;

Alignment Scores:
Pred. No.: 0.00331 Length: 676
Score: 191.00 Matches: 146
Percent Similarity: 37.16% Conservative: 71
Best Local Similarity: 25.00% Mismatches: 221
Query Match: 4.71% Indels: 146
DB: 1 Gaps: 30

US-09-914-958b-35 (1-2177) x MUC1_MESAU (1-676)
QY 229 CTGCACTCACCACCTCAGTTCCTGCTGCTGACTATGACAGTAATCCACCCAGCTCAACT 288
Db 5 ileArgAlaProPheLeuLeuThrLeuLeuLeuAlaLeuValThrAspProAsnSerVal 24
QY 289 ATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAGCTTTGAGATCGTCGTGGGGGGT 348
Db 25 AlaLeuSerGln-----AspThrSerSerSerSerThrLeuAsn 37
QY 349 ACAATGGACCCACCTGTTCTGAAGATGATACCGGGCCCGAGAGCGGCTGGTGAACATCG 408
Db 38 Thr-----ThrProValHisSerGlySerAlaPro 48
QY 409 CGGTGATGAGCGCAGCTCACCTACTACGCGTGGGG---ACCGGAGGGGAGACGCCA 465
Db 49 AlaThrSerSerAlaValAspSerAlaThrThrProGlyHisSerGlySerSerAlaPro 68
QY 466 TTGGGGTCCAGCCTCGCACATCGACGGGAGCGGGAGGAGATCTACTTCTCTCAACA 525
Db 69 ProThrSerSerAlaValAsnSerAlaThrThrProGlyHisSerGlySerSerAlaPro 88
QY 526 CCAATAATGCCCTTCTCGGGGTGGCCACGTACACCGACAGTGTTCAGTTCGCGCAATA 585
Db 89 ProThrSerSerAlaValAsnSerAlaThrThrProValHisSerGlySerSerAlaPro 108
QY 586 ACGGTGGGAGACATCTGAGCGATGAGTCA---ACGTGGCCCTGTGTGGCCAGCC 642
Db 109 Val-----ThrSerSerAlaValAsnSerAlaThrThrProValHisSerGlySer 125
QY 643 TCTTTTCCGAGCCTCTGTGGCCTGTGTGGACAAAAGGGCTCTGGACGCTACTCTA--- 699
Db 126 SerAlaProPro-----ThrSerSerAlaValAsnSerAlaThrThr 139
QY 700 -----TCTACATTCGCAATTCACCTACG---CTAATGTGGGCGCTGATG 741
Db 140 ProValHisSerGlySerSerAlaProValThrSerSerAlaValAsnSerAlaThrThr 159
QY 742 CCCTCATTGAATGGACCTCGAGGCGCAGTGCATCTCTCCCGGGGCAITTCGCGCTCAGAG 801
Db 160 ProValHisSerGlySerSerAlaProValThrSerSerAla----- 173

```

QY 802 ATGTGGCTGCTAGGCTGGGTCAGCAAAATATACAGGGGGCCGAGGCGTCAAGCTGGGCC 861  
Db 174 -----ValAspSerAlaThrThrProValHisSerGlySerSerAlaPro 188  
QY 862 CCATCTCTCAGCAGCAGTCCGTCGATATCTTCTCGACAAATGAGAAATGGCGCTAACTTCC 921  
Db 189 ProThrSerSerAlaVal-----AsnSerAlaThrThrProVal----- 201  
QY 922 TTTTCCACACCGGGCGATGGACCTTTGTGGACGCTGGCGGCAGTCTGGTGTGGACG 981  
Db 202 -----HisSerGlySerSerAlaProVal-----ThrSerSerAlaValAsnSerAlaThr 218  
QY 982 ACCCCACACGATGGCGAGGTCTCGCCTGGTCTGACTTCAACCGTGATGGCAAGTGG 1041  
Db 219 ThrProValHisSerGlySerSerAlaPro-----ValThrSerSerAlaValAsnSerAla 237  
QY 1042 ACATGCTTATGGCAACTGGAATGGCC-----CCACCGGCTCTATCTGCMAATGACCA 1095  
Db 238 ThrThrProValHisSerGlySerSerAlaProProThrSerSerValValAsn----- 255  
QY 1096 CCCATGGGAAGTCCGCTTCGGGACATCGCCT---CACCAAGTCTCCATGCCCTCC 1152  
Db 256 -----SerAlaThrThrProValHisSerGlySerSerAlaProPro 269  
QY 1153 CTGTCCGCAAG-----TCATCACCGCCGACTTGTACATGACACGAGCTGGAGA 1203  
Db 270 ThrSerSerAlaValAsnLeuAlaThrThrProValHisSerGlySerSerThrProAla 289  
QY 1204 TCTTCTTCAACACATTCGCTACCGAGCTCCCTCAGCCACCGCTCTTCGGCGTCTATCC 1263  
Db 290 ThrAsnSerThrThrAspSerAlaThrThrProValPro-----ProGlySerSer 306  
QY 1264 GTAGAGACGACGAGACCCCTCATCAGAGAGCTCAATCCCGCGAGCGCTTGGAGCCTG 1323  
Db 307 MetGlnThrThrGluAlaThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 326  
QY 1324 AGGGCCGGGACAGCGGGTGTGGTACGACACTTCGACG-----GAGACGGGATGCTGG 1377  
Db 327 ValProThrThrSerSerAlaLeuValProThrThrSerAlaAlaHisSerGlyAlaSer 346  
QY 1378 ACCTCATCTTGTCCCATGGAGTCCATGCTCAGCGCTGCTCCGCTTTCGGGGCAATC 1437  
Db 347 AlaMetThrAsnSerSerGluSerAsp-----LeuAlaThrThrProLeuAspSerGlyThr 365  
QY 1438 AGGCTTCAACACAACTGCTCGAGTGTGGTGCACGACCGGTTGGGGCCTTTGCCA 1497  
Db 366 SerIleSerThrThr-----LysAlaPro----- 373  
QY 1498 GGGAGCTAAGTGTGCTCTACACCAAGAAGTGGG-----CCCACTGAGGA 1548  
Db 374 -----AlaThrThrProValHisAsnGlySerLeuValProThrThrSer 388  
QY 1549 TCATCGACGGGGCTCAGGCTACCTGTGTCAGATGAGCGCGTGGCAGACATTTGGCGCTG 1608  
Db 389 SerValLeuGlySerAlaThrThrLeuIleHis-----AsnAspThrSerThrMet----- 405  
QY 1609 GGAAGATGAAGCCAGCTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGGAGCGGA 1668  
Db 406 -----AlaThr 407  
QY 1669 AGTGGCCAGCGGGAGATGAATCAGTCTGAGATCTCTTACCCCGGGATGAGGACA 1728  
Db 408 ThrThrProValGlyAsnGlyThrGln-----SerSerValProSer----- 421  
QY 1729 CACTTCAGGACCCGCCCACTGAGTGTGGCGCAAGGATTCCTCCAGCAGGAAATGGCC 1788  
Db 422 -----ArgHisPro 424  
QY 1789 ATGTGATGACACCAATGAATGATCCAGTCCATTCGTGTCCTCGAGCAAGACCGCG 1848  
Db 425 ValThrProThrProAlaValSerSerAsnSerThrIleAlaLeuSerThr----- 442  
QY 1849 TATGTGTCAACA-----CCTATGGAGCTACAGGTGCGGACCAACAAGAGTGCA 1899

Db 443 TyrThrSerThrAlaLeuSerProAlaPheSerHisAlaAlaPro---GlnValSer 461  
QY 1900 GTCGGGCTAGC 1911  
Db 462 ValGlyValSer 465  
RESULT 13  
DRPL HUMAN  
ID DRPL HUMAN STANDARD; PRT; 1185 AA.  
AC P54259; Q99495; Q99621; Q9UEK7;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 10-OCT-2003 (Rel. 42, last sequence update)  
DT 10-OCT-2003 (Rel. 42, last annotation update)  
DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).  
GN DRPLA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain, and Cerebellum;  
RX MEDLINE=95144175; PubMed=7842016;  
RA Nagatuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,  
Inoue T., Yamada M.;  
RT "Structure and expression of the gene responsible for the triplet  
repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";  
RL Nat. Genet. 8:177-182(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=96026098; PubMed=7485154;  
RA Onodera O., Oyake M., Takano H., Ikeuchi T., Igarashi S., Tsuji S.;  
RT "Molecular cloning of a full-length cDNA for dentatorubral-  
pallidolysian atrophy and regional expressions of the expand alleles  
in the CNS.";  
RL Am. J. Hum. Genet. 57:1050-1060(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96262314; PubMed=8956562;  
RA Margolis R.D., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,  
Kidwai A.S., Ashworth R.G., Ross C.A.;  
RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human  
brain.";  
RL Brain Res. Mol. Brain Res. 36:219-226(1996).  
RN [4]  
RP SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.  
RC TISSUE=Brain;  
RX MEDLINE=97228904; PubMed=9074930;  
RA Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;  
RT "Large-scale sequencing in human chromosome 12p13: experimental and  
computational gene structure determination.";  
RL Genome Res. 7:268-280(1997).  
RN [5]  
RP SEQUENCE OF 1-76 FROM N.A.  
RX MEDLINE=97005364; PubMed=8852663;  
RA Yanagisawa H., Fujii K., Nagafuchi S., Nakahori Y., Nakagome Y.,  
Akane A., Nakamura M., Sano A., Komure O., Kondo I., Jin D.K.,  
Soerensen S.A., Potter N.T., Young S.R., Nakamura K., Nukina N.,  
Nagao Y., Tadokoro K., Okuyama T., Miyashita T., Inoue T.,  
Kanazawa I., Yamada M.;  
RT "A unique origin and multistep process for the generation of expanded  
DRPLA triplet repeats.";  
RL Hum. Mol. Genet. 5:373-379(1996).  
RN [6]  
RP SEQUENCE OF 470-725 FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=93315145; PubMed=8325628;  
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;  
RT "Novel triplet repeat containing genes in human brain: cloning,  
expression, and length polymorphisms.";  
RL Genomics 16:572-579(1993).

[7] INTERACTION WITH WWP1 AND WWP2.  
 RP MEDLINE=98313405; PubMed=9647693;  
 RA Wood J.D., Yuan J., Margolis R.L., Colomer V., Duan K., Kushi J.,  
 RA Kaminsky Z., Kleiderlein J.J. Jr., Sharp A.H., Ross C.A.;  
 RA "Atrophin-1, the DRPLA gene product, interacts with two families of  
 RT WW domain-containing proteins.";  
 RL Mol. Cell. Neurosci. 11:149-160(1998).  
 CC -!- SUBUNIT: Interacts with WWP1 and WWP2.  
 CC -!- TISSUE SPECIFICITY: Relatively high levels in the brain, ovary,  
 CC testis and prostate. Lower levels in the liver, thymus and  
 CC leukocytes.  
 CC -!- POLYMORPHISM: The poly-Gln region of DRPLA is highly polymorphic  
 CC (7 to 23 repeats) in the normal population and is expanded to  
 CC about 49-75 repeats in DRPLA patients. Longer expansions result in  
 CC earlier onset and more severe clinical manifestations of the  
 CC disease.  
 CC -!- DISEASE: Defects in DRPLA are the cause of dentatorubral-  
 CC pallidolysian atrophy (DRPLA) [MIM:125370], an autosomal dominant  
 CC neurodegenerative disorder characterized by a loss of neurons in  
 CC the dentate nucleus, rubrum, globus pallidus and lysis/body.  
 CC Clinical features are myoclonus epilepsy, dementia, and cerebellar  
 CC ataxia. Onset of the disease occurs usually in the second decade  
 CC of life and death in the fourth.  
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to several  
 CC frameshifts.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D31840; BAA06626.1; -.  
 CC DR EMBL; D38529; BAA07534.1; ALT\_FRAME.  
 CC DR EMBL; U23851; AAB50276.1; -.  
 CC DR EMBL; U47924; AAB51321.1; -.  
 CC DR EMBL; D63808; BAA23631.1; -.  
 CC DR EMBL; L10377; -; NOT\_ANNOTATED\_CDS.  
 CC DR PIR; G01763; G01763.  
 CC DR Genew; HGNC:3033; DRPLA.  
 CC DR MIM; 607462; -.  
 CC DR MIM; 125370; -.  
 CC DR GO; GO:0005737; C:cytoplasm; TAS.  
 CC DR GO; GO:0005634; C:nucleus; TAS.  
 CC DR GO; GO:000515; F:protein binding; IPI.  
 CC DR GO; GO:0007417; P:central nervous system development; TAS.  
 CC DR InterPro; IPR002951; Atrophin.  
 CC DR Pfam; PF03154; Atrophin-1; 2.  
 CC DR PRINTS; PR01222; ATROPHIN.  
 CC KW Triplet repeat expansion; Polymorphism; Epilepsy.  
 CC FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).  
 CC FT DOMAIN 302 305 POLY-PRO.  
 CC FT DOMAIN 376 382 POLY-SER.  
 CC FT DOMAIN 386 397 POLY-SER.  
 CC FT DOMAIN 442 447 POLY-PRO.  
 CC FT DOMAIN 479 483 POLY-HIS.  
 CC FT DOMAIN 484 497 POLY-GLN.  
 CC FT DOMAIN 504 507 POLY-PRO.  
 CC FT DOMAIN 564 574 POLY-SER.  
 CC FT DOMAIN 704 707 POLY-PRO.  
 CC FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).  
 CC FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).  
 CC FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).  
 CC FT DOMAIN 94 94 MISSING (IN REF. 3).  
 CC FT CONFLICT 333 333 H -> Y (IN REF. 1).  
 CC FT CONFLICT 339 339 M -> I (IN REF. 3).  
 CC FT CONFLICT 541 541 P -> T (IN REF. 6).  
 CC FT CONFLICT 1028 1028 A -> G (IN REF. 1).  
 CC SEQUENCE 1185 AA; 124773 MW; 5135744CEE491C18 CRC64;

Alignment Scores: 0.00411 Length: 1185  
 Pred. No.: 190.50 Matches: 165  
 Score: 32.63% Conservaive: 52  
 Best Local Similarity: 24.81% Mismatches: 231  
 Query Match: 4.61% Indels: 218  
 DB: 1 Gaps: 37

US-09-914-958B-35 (1-2177) x DRPLA\_HUMAN (1-1185)

QY 1779 CCTGTGGGAGATCTCTGGCCACACTCCAGTGGGGTGGTCTGGAAGTGTGTCTCAT 1720  
 Db 201 ProThrSerArgMetPhe---GlnAlaProProGlyAlaProPro-----ProHis 216  
 QY 1719 -----CCGGGGGTAGAGGATCTCCAGCACTAGTTCATCTCCCGCTGGGCA 1672  
 Db 217 ProGlnLeuTyrProGlyGlyThrGlyGly---ValLeuSerGlyProProMetGlyPro 235  
 QY 1671 CGTTCGGCTCACCATCTTGGCCATCTGCCACGTCACTCCACACACTGTGGCTTCTCT 1612  
 Db 236 LysGlyGlyGlyAlaAlaSerSerValGly---GlyProAsnGlyGlyLysGlnHisPro 254  
 QY 1611 TCCCAGGCCAAAGTGTCCACGGCTCCATCTCACAGGTAGCTGAGCCCCCGTCGA 1552  
 Db 255 ProPro-----ThrThrProIleSerValSerSerGly----- 266  
 QY 1551 TGATCTCAGTGGGGGGCCCACTCTTCTTGGTGTAGAGCAGCACTTAGCTCCCTGGCAA 1492  
 Db 267 ---AlaSerGlyAlaPro-----ProThrLys 274  
 QY 1491 AGGCCCCAAACGGGTGCTGGCCACCACTCGCAGCAGTGTGTGAAGCCCTGATTC 1432  
 Db 275 ProProThrThr-----ProValGly 281  
 QY 1431 CCGGAAGACGACACGGCTGAGCCATGAGTCTCCATGGGCAAGATGAGTCCAGCA 1372  
 Db 282 GlyGlyAsnLeuProSerAlaPro-----ProAla 292  
 QY 1371 TCCCGTCTCCGTGGAAGTCGGTCAACACAC-----CCCTGTGCCCCCGCTCAGGCT 1318  
 Db 293 Asn-----PheProHisValThrProAsnLeuProProProAla 306  
 QY 1317 CCAAGCGCTGCGCGGATTGAGTCTCTCGATAGGGGGTCTCGTGTCTCTTACGGATGA 1258  
 Db 307 LeuArgProLeuAsnAsnAlaSerAlaSerProProGlyLeuGlyAla----- 322  
 QY 1257 CCGGAAGAGCGGTGGTGGTAGGAGCTCGGTAGCAATGTTGTGAAGAGATCTCCA 1198  
 Db 323 -----GlnPro 324  
 QY 1197 GCTCTGTGTCATTGTCAAGTCGGCGGTGATGACCGTG-CGGACAGGGAGGATGGAG 1139  
 Db 325 LeuProGlyHisLeuProSer-----ProHisAlaMetGlyGlnGlyMetGly 340  
 QY 1138 AACTTGGGTGAGCGGATGTCCTCGGAGAGCGAOCCTTCCCATGGTGTCTATTTCAGATAG 1079  
 Db 341 GlyLeuProProGlyProGlyLys----- 348  
 QY 1078 AGCGGTGGGGCCATTCAGTTGCCATAGACGATCCACTTCCATCCATCAGCGTTGAG 1019  
 Db 349 -----GlyProThrLeuAlaPro---SerProHisSerLeuProProAlaSerSer 364  
 QY 1018 TCAGCCAGGGCGACACTCGCCCATCTGCTGGTGGGGTCCGTCCACACAGCATGGCGCA 959  
 Db 365 SerAlaProAlaProProMetArgPheProTyrSerSerSerSerSerSerAlaAla 384  
 QY 958 GCGTCCACAAAGTGCATCGCCCCGGTGTGGAAAGAGTAGGCCCATTTCTCATTTG 899  
 Db 385 AlaSerSerSerSerSerSerSer-----SerSerSerAlaSerProProAla 401  
 QY 898 TCGCAGAGATATCCGAGGCACTGCTGTGAGGATGGGGCCCGCTGAGCGCTCGGCC 839  
 Db 402 SerGlnAlaLeuProSerTyr-----ProHisSerProPro---Pro 414

```

QY 838 CCTGTATATTGCTACCCAGCCTCAGCAGCCACATCTCTGAGCGCCAGCAATGCCCGG 779
Db 415 ProThrSerLeuSerVal-SerAsnGlnProProLysTyrThrGlnProSerLeuProSe 434
QY 778 GAGAGGTCACTGGCCTCAGGGTCCA-----TT 752
Db 434 :GlnAlaValTipSerGlnGlyProProProProProTyroTyrGlyArgLeuLeuAlaAs 454
QY 751 TCAATGAGGCGCATCAGGCGCCACATTACCGTAGGCGTAATGGCAATGTAGATAGTAG 692
Db 454 nSerAsnAlaHisProGlyProPheProProSerThrGlyAlaGln-----SerTh 471
QY 691 CGTCCAGAGCCCTTCTGTGCACAC-----AGGCCACAGACGCT 653
Db 471 xAlaHisProProValSerThrHisHisHisHisGlnGlnGlnGlnGlnGln 491
QY 652 CCGGCAAGAGGTGGCCACACACCGGCGCCACCGTGTGACCTCATCGCTCAGGATGCTCTCC 593
Db 491 nGlnGlnGlnGlnGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHis 506
QY 592 CACCGGTATTGCGGAACCTTGAACTTCTCGGTGACGTGGCCACCCCGGAGAGGCA 533
Db 506 oProGly-----AlaPheProHisProLeuGlu-- 515
QY 532 TTATTGGTGTGAGGAAGTAGATCTCTCCCGCGCGTCCCGTCCGATGTCGCGAGGTGTG 473
Db 516 -----GlyGlySerSerHisHisAlaHisPro-----TyrAlaMetSe 528
QY 472 ACCCCAATGCGTTC-----CCTGCGGTGCCGCA-----GCGCG 437
Db 528 rProSerLeuGlySerLeuArgProTyroProProGlyProAlaHisLeuProProProHi 548
QY 436 TAGTAGGTGACGTGCGCT-----CATCGACCGCGATGTCCACCGCCTTC 389
Db 548 sSerGlnValSerTyroSerGlnAlaGlyProAsnGlyProProValSerSerSerSerAs 568
QY 388 TGGGCGCGGTATCTACTACGACACCGAGTTGGTCCATTCTACCCCGCAGGAGTCTCA 329
Db 568 nSerSerSerSerSerGln-----GlySerTyroCysSerHisProSerProSerGI 587
QY 328 AAGTCCCATCATGTTCCACAT-----CAGTAATCTGCCACACCATAGTTGAGTGGGTG 275
Db 587 nGlyProGlnGlyAlaProTyroProPheProProValPro-ThrValThrThrSerSerA 607
QY 274 GGATT-----ACTGTCATAGTCAGGAGCAGACATGAGTTGGTGACTGACGTGACATG 221
Db 607 laThrLeuSerThrValIleAlaThrValAlaSer-----SerProAlaG 622
QY 220 GGTTC-----AGCCCGCTGGGACCCCTCAGTGCATGGCGGAGAA-----183
Db 622 lyTyroLysThrAlaSerProProGlyProProProTyroGlyLysArgAlaProSerProG 642
QY 182 -----CCAGAGCAGCAGCAGG 167
Db 642 lyAlaTyroLysThrAlaThrProProGlyTyroLysProGlySerProProSerPheArgT 662
QY 166 AACGTAATACCTCGCATCGCGGGTCAAGGCTCGGAGCAATCTCCCGCTCTCGGCC 107
Db 662 hrGlyThrProProGlyTyroArgGlyThrSerProProAlaGlyProGlyThrPheLysP 682
QY 106 CCGCGCTAGGGCGTGGGAAGCGGGCGCTCGCTGCCCGCTCTCCCGCGCG-----53
Db 682 roGlySer-ProThrValGlyProGly-----ProLeuProProAlaGlyPro 697
QY 52 -----CCGCGCTGCTGCTCCAGCCCGCTCCGCGGCTGC 14
Db 698 SerGlyLeuProSerLeuProProProAla-AlaProAlaSerGlyProProLeuSe 717
QY 13 CTCGAGC 7
Db 717 rAlaThr 719

```

```

RESULT 14
YMF6 YEAST
ID YMF6 YEAST STANDARD; PRT; 1140 AA.
AC Q04893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN YMR317W OR YM9924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churche C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93 (1997).
CC -!- DOMAIN: Contains many Ser/Thr-rich domain and repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; Z54141; CA930835.1; -.
DR GenBank; U00096.1; YMR317W.
DR SGI; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Alignment Scores:
Pred. No.: 0.00672 Length: 1140
Score: 186.50 Matches: 144
Percent Similarity: 39.38% Conservative: 86
Best Local Similarity: 24.66% Mismatches: 242
Query Match: 4.60% Indels: 113
DB: Gaps: 25

US-09-914-958B-35 (1-2177) x YMF6 YEAST (1-1140)
QY 403 ACATCGCGGTGATGAGCGCAGCTCACCTACTACGCGTGGGACCGGCGGAGGAGC 462
Db 140 SerSerSerAlaIleGlnThrSerSerProGlnThrSerSerSerAsnGlyGlySer 159
QY 463 -----CCATGGGGTTCACAG-----CCTGCACATCGACGGGAGC 498
Db 160 SerSerGluProLeuGlyLysSerSerValLeuGluThrThrAlaSerSerSerAspThr 179
QY 499 GCCGGGAGGATCTACTCTCTCAACACCAATATGCTTCTCGGGGTGGCCACGTACA 558
Db 180 ThrAlaValThrSerSerThrPheThrThrLeuThrAspValSerSerSerProLys-- 198
QY 559 CCGACACAGTCTTCAAGTTCGCAATACCGGT---GGAGAGACATCTGAGCGATGAGG 615
Db 199 ---IleSerSerSerGlySerAlaValThrSerValGlyThrThrSerAspAlaSerLys 217
QY 616 TCAACGTGGCGCGTGTGGCCAGCTCTTTCGCGGAC-----GCTCTGTGGCGCTGTG 669
Db 218 Glu-----ValPheSerSerSerThrSerAspValSerSerSerLeuLeuSerSer 233
QY 670 TGGACAGAAAGGCTCTGGAGCGTACTATCTCATTTGCCAATTCACGCTACGCTAATG 729
Db 234 ThrSerSerProAlaSerSerThrIleSerGluThrLeuProPheSerSerThrIle--- 252

```

1680 GGCGAGATGAACCTCAGTGTGGAGATCCTTACCCCGGGATGAGGACACACTTCAGGAC 1739  
 589 SerSerAla  
 1740 CAGCCCCCAGTGTGGCCCAAGGATTCCTCCAGCAGGAAATGCGCATTCGATGAC 1799  
 601 SerAlaSerLeu-leValThrLysThrLysAsnSerValValSerSerLeuValSerSe 620  
 1800 ACCAATGAATGATCCAGTTCCCATTCGTGTGGCTCTGCGACAGCCCGTATGTGTCAAC 1859  
 620 rIleThrSerSerGluThrThrAsnGluSerAsnLeuAlaThrSer- -SerTh 637  
 1860 ACCTATG- - - - -GAAGCTACAGGTGC 1880  
 637 rSerLeuLeuSerAsnLysAlaThrAlaAgsrLeuSerThrSerAsnAlaThrSerAl 657  
 1881 CGGA- - - - -CCAACAAGAGTGCAGTGGGCTTACGAGCCCAAGGAGGTGGCAGCAGCC 1934  
 657 aSerAsnValProThrGlyThrPheSerSerMetSerSerHisThrSerValIleThrPr 677  
 1935 TGCCTGGGCT 1944  
 677 oGlyPheSer 680  
 RESULT 15  
 PYGO DROME  
 ID PYGO DROME STANDARD; PRT; 815 AA.  
 AC Q9V9W8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Pygopus protein (Gammy legs protein).  
 GN PYGO OR GAM OR CG11518.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21952490; PubMed=11955446;  
 RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
 RT "wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
 RT pygopus to the nuclear beta-catenin-TCF complex."  
 RL Cell 109:47-60 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Head;  
 RX MEDLINE=22010046; PubMed=12015286;  
 RA Parker D.S., Jemison J., Cadigan K.M.;  
 RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling  
 RT in Drosophila";  
 RL Development 129:2565-2576 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherzer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Brottier P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadenport L.B., Davies P.,  
 RA Cherty J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

730 TGGGCGCTGATGCTCATTGAATGAGACCTGAGGCCAGTACCTCT- - - - -CCCGGG 783  
 253 - - - - -LeuSerIleThrSerSerProValSerGluAlaProSer 266  
 784 GCA- - - - -TTCTGGCGCTCA 798  
 267 AlaThrSerSerValSerSerGluAlaSerSerThrSerSerValSerSer 286  
 799 GAGATGTGGCTGTGGCTGAGCAATATACAGGGGGCCGAGGCG- - - - - 849  
 287 Glu- - - - -AlaProLeuAlaThrSerSerValSerSerGluAlaProSerSer 303  
 850 - - - - -TCAGCGTGGGCCCCATCTTCAGCAGCAGTGCCTCGGATCTTCT 894  
 304 ThrSerSerValSerSerGluAlaProSerSerThrSer- - - - -SerSer 319  
 895 CGCAGATGAGATGGCTTCTTCTTCCCAACCGGGGGGATGGACCTTTGG 954  
 320 ValSerSerGluIleSerSerThrSerSerValSerSerGluAlaProLeuAla 339  
 955 ACCTGTGGCGCAGTGTGGACGACCCACGATGGCGGCGATGTCGCCCTGG 1014  
 340 ThrSerSerValSerSerGluAlaProSerSerThrSerSerValSerSerGlu 359  
 1015 CTGACTTCAACCGTATGCGAAAGTGGACATCTGTATGGCACTGGCAATGCCCCACC 1074  
 360 IleSerSerThr- - - - -ThrSerSerValSerSerGluAlaProLeu 374  
 1075 GCCTCT- - - - -ATCGCAATGAGCAGCCCATGGAGTCCGCTTCGGGACATCG 1125  
 375 AlaThrSerSerValSerSerGluAlaProSerSerThrSerSerValSerSer 394  
 1126 CCTCACCAAGTCTCCATGCTCCCTGTCGACGCTATACCCCGCATCCCGATTGACA 1185  
 395 GluAlaProSerSerThrSerSerValSer- - - - -SerGluAlaProSerSerThr 412  
 1186 ATGACGAGGAGTGGAGATCTTCTTCAACAACATGTCCTACCGCAGCTCTCAGCCAC 1245  
 413 SerSerSerValSerSerGluIleSerSerThrLysSerValMetSerSerGluVal 432  
 1246 GCCTTTCGCGCTATCCGTAGAGACGAGACCCCTCTATCAGAGAGCTCAATCCCG 1305  
 433 SerSerAlaThrSerSerSerValSerSerGluAlaProSerSerAlaIleSerSerLeuAla 452  
 1306 GCGAGCGCTTGGAGCTGAGGCGCGGCGCACAGGGGGTGTGTGACCGCATTCGAGCGAG 1365  
 453 SerSerArgLeuPheSerSerLysAsnThrSerValThrSerThrLeuValAlaThrGlu 472  
 1366 ACGGATGTGAGCTCATCTTGTCCATGGAGAGTCCATGCTCAGCGCTGCTCGTCT 1425  
 473 AlaSerSerValThrSerSerLeuArgProSerSerGluThrLeuAla- - - - -Ser 489  
 1426 TCGGGGCAATCAGGCTTCAACAACACTGGCTG- - - - -CGAGTGTGGCCAGCAGC 1478  
 490 AsnSerIleLeuSerSerLeuSerThrGlyThrAsnSerThrValSerThrThrThr 509  
 1479 CGGTTTGGGCGCTTCCCGAGGAGCTAAGTCTGTG- - - - -CTCTACAC 1523  
 510 SerAlaAlaSerSerThrLeuGlySerLysValSerSerSerAsnSerArgMetAlaThr 529  
 1524 AGAAGAGTGGG- - - - -GCCCACTGAGGATCATCGACGGGGCTCA 1565  
 530 SerLysSerSerSerThrSerSerAspLeuSerLysSerSerValIlePheGlyAsnSer 549  
 1566 GGCTACTGTGTGAGATGGAGCCGTGGCACACTTGTGGCTGGGAGGATGAAGCCAGC 1625  
 550 SerThrVal- - - - -ThrThrSerProSerProSerAlaSerIleSerLeuThrAlaSerProLeuPro 568  
 1626 AGTGTG- - - - -GAGGTGACGTGCCAGATGGCAAGATGTGTGACCGGACGTGGCCAGC 1679  
 569 SerValTrpSerAspIleThrSerSerGluAlaSerSerIleSerSerAsnLeuAlaSer 588

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel M.E., Kodira C.D., Kraft C., Kravitz S., Kulp D.A., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 [4]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=Berkely; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RA "A *Drosophila* full-length cDNA resource.";  
 RT Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -I- FUNCTION: Involved in signal transduction through the Wnt pathway.  
 CC -I- SUBUNIT: Binds to RCL9 via the PHD-type zinc finger motif, and  
 CC thereby becomes part of the nuclear ARM/PAN complex.  
 CC -I- SUBCELLULAR LOCATION: Nuclear.  
 CC -I- TISSUE SPECIFICITY: Ubiquitous throughout embryogenesis and larval  
 CC development.  
 CC -I- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
 CC throughout development.  
 CC -I- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF457206; AAL91369.1; -  
 CC EMBL; AF075095; AAL79357.1; -  
 CC EMBL; AF003778; AAF57161.1; -  
 CC EMBL; AF058500; AAL13729.1; -  
 CC FlyBase; FBgn0043900; PYGO.  
 CC GO; GO:0005634; C:nucleus; NAS.  
 CC GO; GO:0030528; F:transcription regulator activity; IPI.  
 CC GO; GO:0030177; P:positive regulation of Wnt receptor signaling; IPI.  
 CC GO; GO:0007367; P:segment polarity determination; IMP.  
 CC GO; GO:0016055; P:Wnt receptor signaling pathway; IGI.  
 CC InterPro; IPR001965; Znf\_PHD.  
 CC Pfam; PF00628; PHD; 1.  
 CC SMART; SM00249; PHD; 1.  
 CC PROSITE; PS01359; ZF\_PHD\_1; 1.  
 CC PROSITE; PS50016; ZF\_PHD\_2; 1.  
 CC Nuclear protein; Trans-acting factor; Wnt signaling pathway; Zinc;  
 KW Metal-binding; Zinc-finger; Segmentation polarity protein.  
 FT DOMAIN 39 45 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT ZN FING 747 805 PHD-TYPE.  
 FT DOMAIN 48 65 ALA-RICH.  
 FT DOMAIN 123 749 ASN/GLY/HIS.  
 FT CONFLICT 393 393 S -> P (IN REF. 1).  
 SQ SEQUENCE 815 AA; 80493 MW; 369FD5A5D34BC136 CRC64;

Alignment Scores: 0.00652 Length: 815  
 Pred. No.: 186.00 Matches: 161  
 Score: 33.64% Conservative: 59  
 Best Local Similarity: 24.62% Mismatches: 222  
 Query Match: 4.59% Indels: 212  
 DB: 1 Gaps: 35  
 US-09-914-958B-35 (1-2177) x PYGO\_DROME (1-815)  
 QY 11 GAGCCAGCCCGGACCGCGGCTGGGA-----GCAAGCAGCGCGCGCGCGCGG 58  
 Db GlnMetGlyProGlyProGlyValGlyLeuProHisMetAsnHis-GlyArgProGl 216  
 QY 59 CGGCAGAGGGGAGCAGCGCGCGCTCCG-----ACGCCCCAGCGCGCGCGCGG 112  
 Db YGlyProGlyProGlyProGlyProValProMetGlySerProMetGlyIleAlaGl 236  
 QY 113 GAGCGGAGGAGTGGCTCCGAGCGCTGACCGCGCATGTCAGGAGTGTACCGTTCTGCT 172  
 Db YMetGlyGlyMetSerProMetGlyGlyMetGlyGlyProSerIleSerProHisHisMe 256  
 QY 173 G---CTGCTCTGGTTCTGCCCATCTAGGGGTCCAGCGGCTCCAGCGGCTGAACCATGTTCA 229  
 Db tGlyMet-GlyGlyLeuSerProMetGlyGlyGlyProAsnGlyProAsnProArgAlaM 276  
 QY 230 TGCAG---TCACCAACTCAGTCTCGCTCTGCTGCTGACTATGACAGTATCCACCCAGCTCAA 286  
 Db etGlnGlySerPro----- 280  
 QY 287 CTATGTTGTGGCAGTACTGATGGACCATGATGGGAGCTTTGAGATCGTGTGGCGGG 346  
 Db ----- 280  
 QY 347 GTAAATGGACCAACTGTTCTGAAATGATAGCGGCGCCAGAGCGGTGTGTAACAT 406  
 Db -----MetGlyGlyProGlyGln-----A 287  
 QY 407 CGCGGTGATGAGCGAGCTACCTTACTACGCGCTGCGGACCGGCGGAGCGGAGCGCAT 466  
 Db snSerProMetAsnSerLeuProMetGlySerProMetGlyAsnProIleGlySerProL 307  
 QY 467 TGGGGTTCACAGCTCGGACATCGACGCGGACGCGCGGAGGAGATCTACTTCTCCACAC 526  
 Db euGlyProSerGlyProGlyPro-GlyAsnProGlyAsnThrGlyGlyProGln--- 325  
 QY 527 CAATAATGCTTCTCGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCGCAATAA 586  
 Db -----GlnGlnGlnGlnGlnProProGlnPro 334  
 QY 587 CGGTGGGAAGACATCTTGAGCGATGAGTTCACGCTGCGGCGGTGGCGCGGTGTCGCGCGCTTT 646  
 Db ProMetAsnAsn-----GlyGlnMetGlyPro-----ProPro--- 345  
 QY 647 TGCCGAGCGCTCTGTGGCCTGTGTGGCAGAAAGGCTCTGGACGCTACTCTATCTACAT 706  
 Db -----LeuHis 347  
 QY 707 TGCCAATTAGCGCTACGGTAATGTGGCGCCTGATGCCCTCATTTGAATGACCGCTGAGGC 766  
 Db SerProLeuGly-----AsnGlyProThrGly 356  
 QY 767 CAG-----TGACCTCTCCGCGGCTCTTGGCGCTCAGAGATGGC 808  
 Db HisGlySerHisMetProGlyGlyProIleProGly---ProGlyProGlyGly 375  
 QY 809 TGCTGAGGCTGGGGTTCAGCAATAATATACAGGCGCGCGCGGTGAGCGGTGCGGCCATCTCT 868  
 Db LeuValGlyProGly-----GlyIleSerProAlaHisGlyAsnAsnPro 390  
 QY 869 CAGCAGCAGTGTCTGGATATCTTCTGGCAGCAATGAGAAATGGCGCTTCTCTTTTCCA 928  
 Db -----



```
Db 391 -GlyGlySerGlyAsnAsnMetLeuGlyGlyAsnProGlyGlyGlyAsn-----SerAs 408
QY 929 CAACGGGGCGATGCGACTTTGTGGACGCTGGCGCAGTGTGCTGTGGAGACCCCA 988
:|||||
Db 408 nAsnAsnGlySerAsnThr-----SerAsnAlaSerAsnAsnGlnAsnProHi 425
QY 989 C-----CAGCATGGCGA--GGTGTGCGCCCTGGCTGACTTCAACCGTGATGGCAA 1036
|
Db 425 sLeuSerProAlaAlaGlyArgLeuGlyVal-Pro----- 436
QY 1037 AGTGACATCGTCTATGCAATGCAATGGCCCGCCACCGCTCTATCTGCAATGAGCAC 1096
|||||
Db 437 -----ThrSerMetGlnSerAsnGly-----ProSerValSer----- 447
QY 1097 CCATGGGAAGTCCGCTTCGGGGAATCGCTCACCACCAAGTTCTCCATGCCCTCCCTGT 1156
|||||
Db 448 -----SerValAlaSerSerValProSerProAlaThr----- 459
QY 1157 CGGCACGGTCATCACCGCGGACTTTGCAATGACACGAGGCTGGAGATCTTCTCAACAA 1216
|||||
Db 460 -----ProThrLeuThrProThrSerThrAlaThrSerMetSerThrS 474
QY 1217 CATTCCTACCGAGCTCCTCAGCCAAACCGCTCTTCGCG-----TCATCCGTAGAGA 1270
:|||||
Db 474 erValProThrSerSerProAlaProProAlaMetSerProHisSerLeuAsnSerA 494
QY 1271 GCACGGAGACCCCTCATCGAGAGCTCAATCCGCGGACGCTTGGAGCCTGAGGGCG 1330
:|||||
Db 494 laGlyProSerProGlyMetProAsnSerGlyPro---SerProLeuGlnSerProAlaG 513
QY 1331 GGGCACAGGGGTGTGGTACCGGACTTCGACGAGACGGGATGCTGGACCTCATCTTGT 1390
|||
Db 513 lyProAsnGly-----ProAsnAsnAsnSerAsn-----AsnAsnAsnGlyP 528
QY 1391 CCATGGAGAGTCCATGGCTCAGCGCTGTCCGTCCTTCGCGGCAATCAGGGCTTCAACAA 1450
|||||
Db 528 roMetMetGlyGlnMetIle-----ProAsnAlaValProMetGlnHisGlnGlnH 545
QY 1451 CAACTGGCTGCAGTGGTCCACGACCC-----GGTTGGGGCTTTGCCAGGG- 1500
|||
Db 545 isMetGly-----GlyGlyProGlyHisGlyProGlyProMetProGlyM 561
QY 1501 -----GAGCTAAGTCTGCTCTACACCAAGAGAGTGGGGCCACCTGAG--- 1546
:|||||
Db 561 etGlyMetAsnGlnMetLeuProProGlnProSerHisLeu-GlyProProHisPro 580
QY 1547 -----GATCATCGAGGGGCTCAGGCTACCTGTGTGA 1579
|||
Db 581 AsnMetMetAsnHisProHisHisProHisHisProGlyGlyProProHisMet 600
QY 1580 GATGAGCCCGTGGCAGACTTTGGCTGGGGAAGGATGAAGCCAGTGTGGAGGTGAC 1639
|||
Db 601 MetGlyGlyProGlyMetHisGlyGlyProAlaGlyMetProProHisMetGlyGlyGly 620
QY 1640 GTGGCCAGATGGCAAGATGGTGAACGCGAAGTGGCGCAGCGGGAGATGAACCTCAGT--- 1696
|||
Db 621 ProAsnProHisMetMetGlyGlyPro---HisGlyAsnAlaGlyProHisMetGlyHis 639
QY 1697 -----GCTGGAGATCCTTACCC---CCGGATGAGGACACACTTCAGACCCAGC 1744
|||
Db 640 GlyHisMetGlyGlyValProGlyProGlyProGlyProGlyProGlyMetAsnGlyProPro 659
QY 1745 CCCACTGGAGTGGCCAGGATTCCTCCAGCAGGAAATGGCCATTCATGTCATGGACACCAA 1804
:|||||
Db 660 HisProHisMetSerProHis-----HisGlyHisPro 670
QY 1805 TGAATGCATCCAGTTCCTCATTCGTGCCCC 1834
|||
Db 671 HisHisHisHisAsnProMetGlyGlyPro 680
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 3, 2004, 19:37:45 ; Search time 109 Seconds  
(without alignments)  
12603.360 Million cell updates/sec

Title: US-09-914-958B-35  
Perfect score: 4056

Sequence: 1 cggagagctcagccagcc.....tctcttgaaaaa 2177

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool -DBV=xlp  
-DB=SPREMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09914958 @CGN 1 1 122 @runat\_03082004.161651.29683 -NCPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

SPREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvirus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3233.5	79.7	661	4 Q9NQ80	Q9nq80 homo sapien

## ALIGNMENTS

RESULT 1

ID	Q9NQ80	PRELIMINARY;	PRT;	661 AA.
AC	Q9NQ80;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	ASPIC precursor.			
GN	ASPI1			

OC Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

[1]  
SEQUENCE FROM N.A.

RP TISSUE=Cartilage;

RA Bolton M.C., Wait R., Saklatvala J.;

RT "Cloning of ASPIC, a novel protein secreted by human normal and osteoarthritic cartilage, identified by 2D electrophoresis and mass spectrometry."

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ276171; CAB98267.1; -

DR GO; GO:0008305; C:integrin complex; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.

DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.

2	3190.5	78.7	653	4	Q9NQ79	Q9nq79 homo sapien
3	3181.5	78.4	637	4	Q8TE52	Q8te52 homo sapien
4	3176.5	78.3	650	4	Q9NQ78	Q9nq78 homo sapien
5	3046	75.1	646	11	Q8R555	Q8r555 mus musculus
6	3035	74.8	646	11	Q8BMF1	Q8bmf1 mus musculus
7	2133.5	52.6	451	4	Q8N4H6	Q8n4h6 homo sapien
8	2035.5	50.2	435	11	Q8R3V8	Q8r3v8 mus musculus
9	1967.5	48.5	418	4	Q9NW46	Q9nw46 homo sapien
10	910	22.4	193	11	Q9QX63	Q9qx63 rattus norv
11	265	6.5	1225	5	Q9VR49	Q9vr49 drosophila
12	264	6.5	1081	16	Q7ULJ0	Q7ulu0 rhodopirell
13	260.5	6.4	1028	16	Q7UP15	Q7up15 rhodopirell
14	255.5	6.3	800	3	Q8TFG4	Q8tf94 schizosacch
15	249	6.1	528	6	Q29071	Q29071 sus scrofa
16	249	6.1	1032	16	Q7UUC1	Q7uuc1 rhodopirell
17	248.5	6.1	615	16	Q7UJS7	Q7ujs7 rhodopirell
18	242	6.0	2232	5	Q8IFX6	Q8ifx6 caenorhabdi
19	240	5.8	1637	6	Q9XSV8	Q9xsv8 bos taurus
20	240	5.8	5146	6	Q8SPM4	Q8spm4 bos taurus
21	238	5.9	1806	5	Q869R4	Q869r4 dictyosteli
22	234	5.8	1052	16	Q7UUB8	Q7uub8 rhodopirell
23	234	5.8	1349	4	Q8WQ4	Q8wwq4 homo sapien
24	231.5	5.7	1237	2	Q7WYN2	Q7wyn2 aceticvibrio
25	226	5.6	1325	5	Q9BKV7	Q9bkv7 leishmania
26	225	5.5	1107	4	Q9H4D6	Q9h4d6 homo sapien
27	224.5	5.5	1157	4	Q9GZM2	Q9gzm2 homo sapien
28	224.5	5.5	1199	4	Q9H487	Q9h487 homo sapien
29	224.5	5.5	1201	4	Q9H486	Q9h486 homo sapien
30	224.5	5.5	1214	4	Q9H485	Q9h485 homo sapien
31	224.5	5.5	1215	4	Q9H484	Q9h484 homo sapien
32	224.5	5.5	1256	4	Q9H483	Q9h483 homo sapien
33	224.5	5.5	1827	4	Q9H482	Q9h482 homo sapien
34	224.5	5.5	2117	4	Q9H4D8	Q9h4d8 homo sapien
35	224.5	5.5	2167	4	Q9H481	Q9h481 homo sapien
36	224.5	5.5	2169	4	Q9NY09	Q9ny09 homo sapien
37	222.5	5.5	1187	4	Q9GZV6	Q9gzv6 homo sapien
38	222	5.5	769	5	Q17921	Q17921 caenorhabdi
39	219.5	5.4	1037	16	Q7UJ36	Q7uuj36 rhodopirell
40	219	5.4	716	4	Q9NYE4	Q9nye4 homo sapien
41	218.5	5.4	944	4	Q9UP82	Q9up82 homo sapien
42	217.5	5.4	1045	4	Q76037	Q76037 homo sapien
43	215.5	5.3	825	5	Q7YZW4	Q7yzw4 caenorhabdi
44	214.5	5.3	814	11	Q8OUR5	Q8our5 mus musculus
45	214.5	5.3	1032	5	P91365	P91365 caenorhabdi

DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 KW SIGNAL. 1 27 POTENTIAL.  
 FT SIGNAL 12 27 POTENTIAL.  
 FT SIGNAL 9 27 POTENTIAL.  
 FT CHAIN 28 661 ASPIC.  
 SQ SEQUENCE 661 AA; 71420 MW; 2BF75321817F0D01 CRC64;

Alignment Scores:  
 Pred. No.: 2,35e-193 Length: 661  
 Score: 3233.50 Matches: 611  
 Percent Similarity: 97.92% Conservative: 2  
 Best Local Similarity: 97.60% Mismatches: 2  
 Query Match: 79.72% Indels: 11  
 DB: 4 Gaps: 1

US-09-914-958B-35 (1-2177) x Q9NQ80 (1-661)

QY 123 ATGGCTCCGAGCGCTGACCCCGCATGTCACGATGTTACCGTTCCTGCTGCTCTGG 182  
 DB 1 MetAlaProSerAlaAspProGlyMetSerArgMetLeuProPheLeuLeuLeuLeuTrp 20  
 QY 183 TTTCTGCCCATCAGTGGGGTCCAGCGGGTGAACCCATGTTCACTGCAGTCAACCAAC 242  
 DB 21 PheLeuProIleThrGluGlySerGlnArgAlaGluProMetPheThrAlaValThrAsn 40  
 QY 243 TCAGTTCCTGCTCCGACTATCACAGTAAATCCACCCAGCTCAACTATGCTGGCAGTT 302  
 DB 41 SerValLeuProAspTrpAspSerAsnProThrGlnLeuAsnTrpGlyValAlaVal 60  
 QY 303 ACTGATGGGACCATGATGGGACTTTGAGATCGTCGTGGCGGGGTACAATCGACCCCAAC 362  
 DB 61 ThrAspValAspHisAspGlyAspPheGluIleValAlaGlyTyrAsnGlyProAsn 80  
 QY 363 CTGGTTCGAAGTATGACCGGCCCCAGAACCGGCTGGTGAACATCGCGTCCATGAGCGC 422  
 DB 81 LeuValLeuLysTyrAspArgAlaGlnLysArgLeuValAsnIleAlaValAspGluArg 100  
 QY 423 AGCTCACCTACTAGCGCTCGGACCGGACCGGAGGGAACCGCATTTGGGGTCAAGCTGC 482  
 DB 101 SerSerProTyrTyrAlaLeuArgAspArgGlnGlyAsnAlaIleGlyValThrAlaCys 120  
 QY 483 GACATCGACGGGCGGCGGAGGAGATCTACTTCTCAACACCAATAATGCTTCTCG 542  
 DB 121 AspileAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAlaPheSer 140  
 QY 543 GGGGTGGCCAGTACACCGACAAGTTGTTCAAGTTCGGAATAACCGGTGGGAAGCATC 602  
 DB 141 GlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAspile 160  
 QY 603 CTGAGCGATGAGTCAACCTGGCGCTGTGTGGCCAGCTCTTTCGCGAGCGCTCTGTG 662  
 DB 161 LeuSerAspGluValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSerVal 180  
 QY 663 GCCTGTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTACATTCGCAATTACGCTTAC 722  
 DB 181 AlaCysValAspArgLysGlySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyr 200  
 QY 723 GGTAAATGGGCCCTGATGCCCTCATTTGAATGGACCTGAGGCCAGTGCACCTCTCCCG 782  
 DB 201 GlyAsnValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSerArg 220  
 QY 783 GGCATTCTGGCGCTCAGAGATGGCTGCTGAGCTGGGGTCAAGCAATATACAGGGGC 842  
 DB 221 GlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrGlyGly 240  
 QY 843 CGAGGGCTCAGCTGGGCCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTCGCAAT 902  
 DB 241 ArgGlyValSerValGlyProIleLeuSerSerAlaSerAspilePheCysAspAsn 260

QY 903 GAGATGGGCTTAACCTTCTTTTCCAAACCGGGCGATGGCACCTTTGTGTGACGCTGG 962  
 DB 261 GluAsnGlyProAsnPheLeuPheHisasnArgGlyAspGlyThrPheValAspAlaAla 280  
 QY 963 GCCAGTGTGTGTGGACGACCCCAACAGCATGGCGAGGTGTGCGCTTGCTGCTACTTC 1022  
 DB 281 AlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPhe 300  
 QY 1023 AACCGTGTGCAAGTGGACATGCTCTATCGCACTGGAATGGCCCCACCGCTCTAT 1082  
 DB 301 AsnArgAspGlyLysValAspileValTyrGlyAsnTrpAsnGlyProHisArgLeuTrp 320  
 QY 1083 CTGCAATGAGCACCCATGGGAAGTCCGCTTCCGGGACATCGCTCACCCAAAGTTCTCC 1142  
 DB 321 LeuGlnMetSerThrHisGlyLysValArgPheArgAspileAlaSerProLysPheSer 340  
 QY 1143 ATGCCCTCCCTGTCGACCGTCAACCGCGCATCTTGTGACATGACGAGAGCTGGAG 1202  
 DB 341 MetProSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGlu 360  
 QY 1203 ATCTTCTTCAACAACATTCGCTACCGCAGCTCCTCAGCCAAACCGCTCTTCCGCTCATC 1262  
 DB 361 IlePhePheAsnAsnIleAlaTyrArgSerSerAlaAsnArgLeuPheArgValIle 380  
 QY 1263 CGTAGAGACACGAGACCCCTCATCGAGGAGCTCAATCCCGCGACGCTTGAGCGCT 1322  
 DB 381 ArgArgGluHisGlyAspProLeuIleGluLeuAsnProGlyAspAlaLeuGluPro 400  
 QY 1323 GAGGCCCGGGCACAGGGGGTGTGTGACCGCTTCCGACGAGAGCGGATGCTGACCTC 1382  
 DB 401 GluGlyArgGlyThrGlyValValThrAspPheAspGlyAspGlyMetLeuAspLeu 420  
 QY 1383 ATCTGCTCCATGAGAGTCCATGCTCAGCCGCTGCTCGTCCGGGGCAATCAGGCG 1442  
 DB 421 IleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGlnGly 440  
 QY 1443 TTCAACAACAACATGGCTGCGAGTGGTCCACACCCCGGTTTGGGGCTTTCCAGGGGA 1502  
 DB 441 PheAsnAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArgGly 460  
 QY 1503 GCTAAGTGTGCTCTTACACCAAGAGTGGGGCCCTGAGGATCATCAGCGGGGC 1562  
 DB 461 AlaLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGlyGly 480  
 QY 1563 TCAGCTACCTGCTGTGAGATGGACCGCTGGGCACACTTTGGCTGGGGAAGATGAAGCC 1622  
 DB 481 SerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyLysAspGluAla 500  
 QY 1623 AGCAGTGTGGAGTACGCTGGCCAGATGGCAAGATGGTGGAGCGGAACTGGCCACGGG 1682  
 DB 501 SerSerValGluValThrTrpProAspGlyLysMetValSerArgAsnValAlaSerGly 520  
 QY 1683 GAGATGAATCAGTGTGGAGATCCTTACCCCGGGATGAGACACACTTACGAGCCCA 1742  
 DB 521 GluMetAsnSerValLeuGluIleLeuTyrProArgAspGluAspThrLeuGlnAspPro 540  
 QY 1743 GCCCCTACCTGCTGGCCCAAGGATTCCTCCAGCAGGAAAATGGCCATTCATGGACACC 1802  
 DB 541 AlaProLeuGluCysGlyGlnGlyPheSerGlnGlnGluAsnGlyHisCysMetAspThr 560  
 QY 1803 AATGAATGATCATCGATTCCCTTCCCTTCGAGCAAGCCCGTATGTGTCAACACC 1862  
 DB 561 AsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsnThr 580  
 QY 1863 TATGGAAGTACAGGTGCGGACCAACAAGAGTGCAGTTCGGGGCTACGAGCCCAACGAG 1922  
 DB 581 TyrGlySerTyrArgCysArgThrAsnLysLysCysSerArgGlyTyrGluProAsnGlu 600  
 QY 1923 GATGGCACAGCTGCTGGCTGGCTGTGTGAAGATAGTGAACACCAAGTT 1982  
 DB 601 AspGlyThrAlaCysValGly-----ThrLeu 609







```
Db 221 AlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSerSer 240
Qy 876 AGTGCTCGGATATCTTCTGGACAAATGAGATGGCCTAACTTCTCTTTTCCACACCGG 935
Db 241 SerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArg 260
Qy 936 GCGGATGCGCCTTTGTGGAGCTCGCGCCAGTGTGGTGGAGCGACCCGCCACGAGAT 995
Db 261 GlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspProHisGlnHis 280
Qy 996 GGGCAGGTGTGCGCCTCGCTGACTTCAACCGTGTATGGCAAAGTGGACATCTCTATGGC 1055
Db 281 GlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyrGly 300
Qy 1056 AACTGGAATGGCCCCACCGCTCTATCTGCAAAATGAGCACCCATGGGAAGTCCGCTTC 1115
Db 301 AsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArgPhe 320
Qy 1116 GCGGATCGCTCACCCCAAGTCTCCATGCCCTCCCTGCTCCGACGCTCATCCGCC 1175
Db 321 ArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThrAla 340
Qy 1176 GACTTTGCAATGACAGAGCTGGAGATCTTCTTCAACAACTTGCCTTACCGAGCTCC 1235
Db 341 AspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSerSer 360
Qy 1236 TCAGCCAAACCGCTCTTCGCTCATCTCGTAGAGACGGAGACCCCTCATTCAGGAG 1295
Db 361 SerAlaAsnArgLeuPheArgValIleArgGluHisGlyAspProLeuIleGluGlu 380
Qy 1296 CTCATTCGGCGAGCCCTTGAGCTGAGGCGCGGCGACAGGGGTGTGTGACCGAC 1355
Db 381 LeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThrAsp 400
Qy 1356 TTCGAGGAGAGCGGATGCTGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCGG 1415
Db 401 PheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGlnPro 420
Qy 1416 CTGTCCGTCTTCGGGGCAATCAGGGCTTCAACAACTGGTGGAGTGGTGGCCACGC 1475
Db 421 LeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValProArg 440
Qy 1476 ACCCGTTTGGGCTTCCAGGGAGCTAGGTGCTGCTTACACCAAGAGAGTGGG 1535
Db 441 ThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSerGly 460
Qy 1536 GCCCACTGAGGATCATCAGCGGGGCTCAGGTACTGTGTGAGATGGAGCCCGTGGCA 1595
Db 461 AlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProValAla 480
Qy 1596 CACTTTGGCTGGGGAAGATGAAGCAGCAGTGTGGAGTGAAGTGGCCAGATGGCAAG 1655
Db 481 HisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGlyLys 500
Qy 1656 ATGCTGAGCGGAGCTGCGCAGCGGGGATCACTGCTGTGAGATGCCCTCTACCCC 1715
Db 501 MetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyrPro 520
Qy 1716 CCGGATGAGACACACTTCAGGACCCAGCCCTGAGTGTGGCCAAAGGATTTCTCCAG 1775
Db 521 ArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGlyPheSerGln 540
Qy 1776 CAGAAATGGCATTTGATGGACCAATGAATGATTCAGTTCCTCCATTCGTGTGCCCT 1835
Db 541 GlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCysPro 560
Qy 1836 CGAGACAAGCCCGTATGTGTCAACCTATGGAAGCTACAGGTGCCGGAACAACAAG 1895
Db 561 ArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLysLys 580
Qy 1896 TGCAGTCGGGGCTACGAGCCCAACGAGATGGCACGCTGCGTGGCGCTGGTGGAGCCCT 1955
Db 581 CysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly----- 596
```

```
Qy 1956 GTGTGTAAGATAGTAGACACCAAGTTGGGAAGAGCCTTGGTCCC 2000
Db 597 -----ThrLeuGlyGlnSerProGlyPro 604
RESULT 5
Q8R555 PRELIMINARY; PRT; 646 AA.
AC Q8R555;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CRTAC1-B protein (ASPIC precursor homolog).
GN 2810454P2IRIK OR CRTAC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN 11 SEQUENCE FROM N.A.
RP STRAIN=BAUB/c; TISSUE=Brain;
RC Steck E., Richter W.;
RA "A novel form of CRTAC1 expressed in human and mouse brain generated
RT by alternative splicing of a newly identified last exon.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN 12 SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Spinal cord;
RC MEDLINB=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AJ421516; CAD13395.1; -;
DR EMBL; AK049801; BAC33924.1; -;
DR MGI; MGI:1920082; Crtac1.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5.
DR PROSITE; PS01187; EGF_CA; 1.
SQ SEQUENCE 646 AA; 70324 MW; 154C45B636DF2D98 CRC64;
Alignment Scores:
Pred. No.: 1.16e-181 Length: 646
Score: 3046.00 Matches: 576
Percent Similarity: 94.89% Conservative: 18
Best Local Similarity: 92.01% Mismatches: 24
Query Match: 75.10% Indels: 8
DB: 11 Gaps: 2
US-09-914-958B-35 (1-2177) x Q8R555 (1-646)
Qy 123 ATGCTCCGAGCGCTGACCCCGCATGTCAGGATGTTACCGTTCCTGCTG---CTGCTC 179
Db 1 MetAlaProSerAlaAspProGlyMetValArgMetAlaLeuLeuLeuProLeu 20
Qy 180 TGTCTTCTGCCCCATCATCTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGAGTCACC 239
Db 21 TrpLeuLeuProLeuThrGlyGlySerGlnArgAlaGluProMetPheThrAlaValThr 40
Qy 240 AACTCAGTCTGCTCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGTGGCA 299
Db 41 AsnSerValLeuProAspTyrAspSerAsnProThrGlnLeuAsnTyrGlyValAla 60
Qy 300 GTTACTGATGTGACCATGATGGGACTTTGAGATCGCTGGCGGGGTACAATGGACCC 359
Db 61 ValThrAspValAspHisAspGlyAspPheGluIleValAlaGlyTyrThrGlyPro 80
```



QY 360 AACCTGGTCTCTGAAGTATGACCGGGCCAGAGCGGCTGTGTAACATCGCGGTGCATGAG 419  
 Db |||||:::|||||  
 QY 81 AsnLeuValLeuLysTyrAsnArgAlaGlnAsnArgLeuValAsnIleAlaValAspGlu 100  
 Db |||||:::|||||  
 QY 420 CGCAGCTCACCCCTACTACTCGGCTCGGGACCGGAGGGAACGCATTTGGGGTCCACAGCC 479  
 Db |||||:::|||||  
 QY 101 ArgSerSerProTyrTyrAlaLeuArgAspArgGlnGlyAsnAlaIleGlyValThrAla 120  
 Db |||||:::|||||  
 QY 480 TGGACATCGACGGGAGCGGCGGGAGAGATCTACTCTCCACACCAATAATGCTTC 539  
 Db |||||:::|||||  
 QY 121 CysAspIleAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAlaPhe 140  
 Db |||||:::|||||  
 QY 540 TCGGGGTGCGCCAGTACCGACCAAGTGTCAAGTTCGCCATAACCGGTGGAGAC 599  
 Db |||||:::|||||  
 QY 141 SerGlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAsp 160  
 Db |||||:::|||||  
 QY 600 ATCTGTAGCGATGAGTCAACGTGGCGCTGGTGTGGCCAGCCCTTTGGCGGAGCTCT 659  
 Db |||||:::|||||  
 QY 161 IleLeuSerAspValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSer 180  
 Db |||||:::|||||  
 QY 660 GTGGCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATGTCGAATACGCC 719  
 Db |||||:::|||||  
 QY 181 ValAlaCysValAspArgThrGlySerGlyArgTyrSerIleTyrIleAlaAsnTyrAla 200  
 Db |||||:::|||||  
 QY 720 TACGGTAATCTGGGCTGTGATGCTCCTCATTTGAATGACACCTGAGGCGACCTCTCC 779  
 Db |||||:::|||||  
 QY 201 TyrGlyAspValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSer 220  
 Db |||||:::|||||  
 QY 780 CGGGCATCTCTGGCGCTCAGAGATGTGGCTGTGAGCTGGGTTCAGCAATATACAGGG 839  
 Db |||||:::|||||  
 QY 221 ArgGlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrAla 240  
 Db |||||:::|||||  
 QY 840 GGGCGAGGCTGAGCTGGGCGCCATCTCTCAGCAGAGTGCCTCGGATATCTCTCGAC 899  
 Db |||||:::|||||  
 QY 241 GlyArgGlyValSerValGlyProIleLeuSerSerSerAlaSerAspIlePheCysasp 260  
 Db |||||:::|||||  
 QY 900 AATGAGAATGGGCTTAACCTCTCTTTCCACACCGGGCGCATGCCACTTTGGAGCT 959  
 Db |||||:::|||||  
 QY 261 AsnGluAsnGlyProAsnPheLeuPheHisAsnGlnGlyAsnGlyThrPheValAspThr 280  
 Db |||||:::|||||  
 QY 960 GGGCGCAGTGTCTGTGGACGACCCCGCCATGCGGCGAGTGTGCGCCCTGGCTGAC 1019  
 Db |||||:::|||||  
 QY 281 AlaAlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAsp 300  
 Db |||||:::|||||  
 QY 1020 TTCACGCTGATGGCAAGTGGACATCGTCTATGGCAACTGGCAATGCGCCCGCCACCGCTC 1079  
 Db |||||:::|||||  
 QY 301 PheAsnArgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeu 320  
 Db |||||:::|||||  
 QY 1080 TATCTCAATGAGCACCCATGGAAGTCCGCTCCGGGACATCGCCTCACCCAGTTC 1139  
 Db |||||:::|||||  
 QY 321 TyrLeuGlnMetSerAlaHisGlyLysValArgPheArgAspIleAlaSerProLysPhe 340  
 Db |||||:::|||||  
 QY 1140 TCCATGCTCCCTCTCCGACGCTCATCCCGCGCATCTTCCACATGACACGAGGAGCTG 1199  
 Db |||||:::|||||  
 QY 341 SerThrProSerProValArgThrValIleAlaAlaAspPheAspAsnAspGlnGluLeu 360  
 Db |||||:::|||||  
 QY 1200 GAGATCTTTTCAACAACATTTGCTACCGAGCTCTCTAGCCAAACCGCTCTTCCCGCTC 1259  
 Db |||||:::|||||  
 QY 361 GluValPhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgVal 380  
 Db |||||:::|||||  
 QY 1260 ATCCGTGAGAGACGACGAGACCCCTCATCGAGAGCTCAATCCCGCGACGCTTGGAG 1319  
 Db |||||:::|||||  
 QY 381 IleArgArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGlu 400  
 Db |||||:::|||||  
 QY 1320 CCTGAGGGCGGGGACAGGGGGTGTGGTACCGACTTCGACGAGACGGATGTGGAC 1379  
 Db |||||:::|||||  
 QY 401 ProGluGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAsp 420  
 Db |||||:::|||||  
 QY 1380 CTCATCTTGTCCCATGGAGATCCATGGTCTGACCGCTGTCCGTCTTCCGGGGCAATCAG 1439  
 Db |||||:::|||||  
 QY 421 LeuIleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGln 440  
 Db |||||:::|||||  
 QY 1440 GGTCTCAACAACATGCTGGAGTGGTGGCAGCACCCGCTTGGGCGCTTGGCCAGG 1499  
 Db |||||:::|||||

Db 441 GlyPheSerAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArg 460  
 QY 1500 GGAGCTAAGGTTCGTGCTCTACACCAAGAGTGGGGCCACCTGAGGATCATCGACGG 1559  
 Db |||||:::|||||  
 QY 461 GlyAlaLysValValLeuTyrThrLysSerGlyAlaHisLeuArgIleIleAspGly 480  
 Db |||||:::|||||  
 QY 1560 GGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCCTGGGAGGATGAA 1619  
 Db |||||:::|||||  
 QY 481 GlySerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyArgAspGlu 500  
 Db |||||:::|||||  
 QY 1620 GCCAGCAGTGTGAGGTGACGTGCCAGATGGCAAGATGGTGGACCCGGAACGTGGCCAGC 1679  
 Db |||||:::|||||  
 QY 501 AlaSerSerValGluValThrTrpProAspGlyLysMetValSerArgSerValAlaSer 520  
 Db |||||:::|||||  
 QY 1680 GGGGAGATGAATCAGTGTGGAGATCTCTACCCCGGGATGAGGACACACTTTCAGGAC 1739  
 Db |||||:::|||||  
 QY 521 GluGluMetAsnSerValLeuGluIleLeuTyrProGlnAspGluAspLysLeuGlnAsn 540  
 Db |||||:::|||||  
 QY 1740 CCAGCCCCACTGGAGTGTGGCCAGGATTTCTCCAGCAGGAGAAATGGCCATTGTCATGGAC 1799  
 Db |||||:::|||||  
 QY 541 ThrAlaProLeuGluCysGlyGlnGlyPheSerGlnGlnAspAsnGlyHisCysMetAsp 560  
 Db |||||:::|||||  
 QY 1800 ACCAATGAATGATCGATCCGTTCCCATTCGTGTGGCTCGACAGACAGCCGCTATGTCTCAAC 1859  
 Db |||||:::|||||  
 QY 561 ThrAsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsn 580  
 Db |||||:::|||||  
 QY 1860 ACCTATGGAGCTACAGGTCCGAGCCGACCAAGAGTGCAGTGGGCTACGAGCCCAAC 1919  
 Db |||||:::|||||  
 QY 581 ThrTyrGlySerTyrArgCysArgThrAsnLysArgCysAsnArgGlyTyrGluProAsn 600  
 Db |||||:::|||||  
 QY 1920 GAGGATGGCACACCTCGCTG-----GGCTGGTGGAGCCCTGTG 1958  
 Db |||||:::|||||  
 QY 601 GluAspGlyThrAlaCysValAlaGlnValAlaPheLeuGlyGlyTyrSerSerAlaAla 620  
 Db |||||:::|||||  
 QY 1959 TTGAAGATGTGACACCA 1976  
 Db |||||:::|||||  
 QY 621 PheArgLeuSerGluPro 626  
 Db |||||:::|||||

RESULT 6  
 Q8BMF1 PRELIMINARY; PRT; 646 AA.  
 AC Q8BMF1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ASPIC precursor homolog.  
 GN CRTL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 SE SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RX MEDLINE=22354683; PubMed=1246851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK032328; BAC27817.1; -.  
 DR MGD; MGI:1920082; Crtac1.  
 DR GO; GO:0008305; C:integrin complex; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 SE SEQUENCE 646 AA; 70197 MW; B8841A37B05FF314 CRC64;

## Alignment Scores:

Pred. No.: 5,61e-181 Length: 646  
Score: 3035.00 Matches: 574  
Percent Similarity: 94.73% Conservative: 19  
Best Local Similarity: 91.69% Mismatches: 25  
Query Match: 74.83% Indels: 8  
Dbs: 11 Gaps: 2

US-09-914-958B-35 (1-2177) x Q8BMF1 (1-646)

```
QY 123 ATGGCTCGAGCGCTACCCCGGATGTCACGAGATGTTACCGTTCTCTGCTG---CTGCTC 179
Db 1 MetAlaProSerAlaAspProGlyMetValArgMetAlaLeuLeuLeuLeuProLeu 20
QY 180 TGGTTTCGCCATCAGTGGGGTCCAGGGGTGAACCCATGTTCACTGCATGACC 239
Db 21 TrpLeuLeuProLeuThrGlyGlySerGlnArgAlaGluProMetPheThrAlaValThr 40
QY 240 AACTCAGTTCCTCCTGACTATGACAGTAAATCCACCCAGCTCAACTATGCTGGCA 299
Db 41 AsnSerValLeuProProAspPyrAspSerAsnProThrGlnLeuAsnTrpGlyValAla 60
QY 300 GTTACTGATGGACCATGATGGGACTTTGAGATCGTCTGGCGGGGTACAAATGGACCC 359
Db 61 ValThrAspValAspHisAspGlyAspPheGluIleValValAlaGlyThrGlyPro 80
QY 360 AACCTGGTTCTGAATATGACCGGGCCAGAGCGGCTGGTGAAATCGCGTCCATGAG 419
Db 81 AsnLeuValLeuLysTyrAsnArgAlaGlnAsnArgLeuValAsnIleAlaValAspGlu 100
QY 420 CGCAGCTCACCTACTAGCGCTGGGACCGGCGGAGGAAACGCAATGGGGTCAAGCC 479
Db 101 ArgSerSerProTyrTyrAlaLeuArgAspArgGGIndlyAsnAlaIleGlyValThrAla 120
QY 480 TCGCATATCGAGCGGACCGCGGAGGAGATCTACTTCTCTCAACACCAATATGCTTC 539
Db 121 CysAspIleAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAlaPhe 140
QY 540 TCGGGGTGGCCAGTACACGACAAAGTTGTCAGTTCCGCAATACCGTGGGAGAC 599
Db 141 SerGlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAsp 160
QY 600 ATCTCGAGCGATGAGTCAACGTCGCGTGTGTGGCCAGCTCTTTCCGCGAGCTCT 659
Db 161 IleLeuSerAspAspValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSer 180
QY 660 GTGGCTGTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTATCTATCCCAATACGCC 719
Db 181 ValAlaCysValAspArgThrGlySerGlyGlyTyrSerIleTyrIleAlaAsnTyrAla 200
QY 720 TACGGTAATGTGGGCCCTGATGCCCTCATTTGAATGGACCCCTGAGCCAGTACCTCTCC 779
Db 201 TyrGlyAspValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSer 220
QY 780 CGGGCATTCGCGCTCAGAGATGTGGCTGTGGCTGGGCTGAGCAATATACAGG 839
Db 221 ArgGlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrAla 240
QY 840 GGCCTGAGGCTCAGCGTGGGCCCCATCTCAGCAGCAGTGCCTCGGATATCTTCTCGAC 899
Db 241 GlyArgGlyValSerValGlyProIleLeuSerSerAlaSerAspIlePheCysAsp 260
QY 900 AATGAGAAATGGGCTTAATCTCTTTTCCACACCGGGCGATGGACCTTTGGAGCGCT 959
Db 261 AsnGluAsnGlyProAsnPheLeuPheHisAsnGlnGlyAsnGlyThrPheValAspThr 280
QY 960 GCGGCGAGTGTGCTGTGACAGACCCACCGACATGGGCGAGCTGTGGCTGGCTGAC 1019
Db 281 AlaAlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAsp 300
QY 1020 TTCAACCGTGATCGCAAAAGTGACATCGTCTATGGCAATCGGCAATGGCCCCACCGCTC 1079
```

```
Db 301 PheAsnArgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeu 320
QY 1080 TATCTGCAATGAGCACCCATGGGAAGTTCGCTTCGCGGACATCGCTCCACCCCAAGTTC 1139
Db 321 TyrLeuGlnMetSerAlaHisGlyLysValGlnPheArgAspIleAlaSerProLysPhe 340
QY 1140 TCCATGCCCTCCCTGTCGCGACCGTTCATCCCGCGACTTTGCAATGACGAGAGCTG 1199
Db 341 SerThrProSerProValArgThrValIleAlaAspPheAspAsnAspGlnGluLeu 360
QY 1200 GAGATCTTCTTCAACAACATTCCTACCGAGCTCTCAGCCACCGCTCTTTCGCGCTC 1259
Db 361 GluValPhePheAsnAsnIleAlaTyrArgSerSerAlaAsnArgLeuPheArgVal 380
QY 1260 ATCCGTGAGAGACGAGACCCCTCATCGAGGAGCTCAATCCCGGCGACCGCTTGGAG 1319
Db 381 IleArgArgGluHisGlyAspProLeuIleGluLeuLeuAsnProGlyAspAlaLeuGlu 400
QY 1320 CCTGAGGCGCGGGCACAGGGGGTGTGTGACCGACTTCGACGGAGACGGGATGTGGAC 1379
Db 401 ProGluGlyArgGlyThrGlyValValThrAspPheAspGlyAspGlyMetLeuAsp 420
QY 1380 CTCATCTTGTCCATGAGAGTCCATGCTCAGCGCTGTCGCTCTTCGGGGGCAATCAG 1439
Db 421 LeuIleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGln 440
QY 1440 GGCTTCAACAACAACACTGGCTGCGAGTGGTCCACGACCCGCTTTGGGCGCTTCCAGG 1499
Db 441 GlyPheSerAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArg 460
QY 1500 GGAGTAAGTGTGCTCTTACACCAAGAGAGTGGGGGCCACTGAGGATCATTCGACGGG 1559
Db 461 GlyAlaLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGly 480
QY 1560 GCCTCAGCTACTCTGTGAGATGGAGCCCGTGGCACACTTTGGCTGGGAGGATGAA 1619
Db 481 GlySerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyArgAspGlu 500
QY 1620 GCACAGTGTGAGAGTACGTGCGGACAGATGGTCAAGATGGTGGAGCCGGAACGTGGCAGC 1679
Db 501 AlaSerSerValGluValThrTrpProAspGlyLysMetValSerArgSerValAlaSer 520
QY 1680 GGGGAGATGAATCTAGTGTGAGATCTCTACCCCGGGGATGAGACACACTTCAGGAC 1739
Db 521 GluGluMetAsnSerValLeuGluIleLeuTyrProGlnAspGluAspLysLeuGlnAsn 540
QY 1740 CCAGCCCCACTCGAGTGTGGCCAGGATCTCCAGCAGGAAAATGGCCATTGTCATGGAC 1799
Db 541 ThrAlaProLeuGluCysGlyGlnGlyPheSerGlnGlnAsnAsnGlyHisCysMetAsp 560
QY 1800 ACCAATGAATGATCATCCAGTTCCCATTCGTGCTGCGCTTCGAGACAAAGCCCGTATGTCAAC 1859
Db 561 ThrAsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsn 580
QY 1860 ACCTATGGAAGTACAGGTGCGCGACCAACAGAGAGTGCAGTGGGGGCTACGAGCCCAAC 1919
Db 581 ThrTyrGlySerTyrArgCysArgThrAsnLysArgCysAsnArgGlyTyrGluProAsn 600
QY 1920 GAGGATGGCACAGCTCGCTG---GGCTGGTGGAGCGCTGCTG 1958
Db 601 GluAspGlyThrAlaCysValAlaGlnValAlaPheLeuGlyTyrSerSerAlaAla 620
QY 1959 TTGAAGATAGTGACACCA 1976
Db 621 PheArgLeuSerGluPro 626
```

## RESULT 7

Q8N4H6 ID Q8N4H6 PRELIMINARY; PRT; 451 AA.  
AC Q8N4H6; 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

Similar to cartilage acidic protein 1.

DE Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=9606;  
 [1]  
 RN PROSITE: PS01187; EGF CA; 1.  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC034245; AAH34245.1; -.  
 DR GO; GO:0008305; C:integrin complex; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR InterPro; IPR001881; EGF\_CA.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR SMART; SM00179; EGF CA; 1.  
 DR PROSITE; PS01187; EGF CA; 1.  
 SQ SEQUENCE 451 AA; 48270 MW; 26F3DE1B9D970D2F CRC64;

Alignment Scores:  
 Pred. No.: 8.5e-125 Length: 451  
 Score: 2133.50 Matches: 401  
 Percent Similarity: 96.88% Conservative: 2  
 Best Local Similarity: 96.39% Mismatches: 2  
 Query Match: 52.60% Indels: 11  
 DB: 4 Gaps: 1

US-09-914-958b-35 (1-2177) x Q8N4H6 (1-451)

QY 753 ATGACCTGAGCGAGTCTCCCGGGGCAATTCGGCGTCAAGATGGCTGCT 812  
 Db 1 MetAspProGluAlaSerAspLeuSerArgGlyLeuAlaLeuArgAspValAlaA 20

QY 813 GAGCTGGGTGAGCAATATACAGGGGCGGAGGCTGAGCTGGGCGCCATCTCAGC 872  
 Db 21 GluAlaGlyValSerLysThrGlyGlyArgGlyValSerValGlyProLeuSer 40

QY 873 AGCAGTGCCTCGGATATCTCTCGCAATAGAAATGGGCTAACTTCCTTTCCCAAC 932  
 Db 41 SerSerAlaSerAspLeuPheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

QY 933 CGGGGCGATGGACCTTTGTGGAGCTGCGGCGGCGAGTGGTGGACGACCCCGCCAG 992  
 Db 61 ArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspAspProHisGln 80

QY 993 CATGGCGAGGTCTCGCCCTGGCTGACTTCAACCGTGTGCGCAAGTGGACATCGTCTAT 1052  
 Db 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspLeuValTyr 100

QY 1053 GGCACTGGAATGGCCCGCCCTCTATCTGCAATAGACACCCATGGGAAGGTCCGC 1112  
 Db 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

QY 1113 TTCCGGGATCGCTCCCGCCATGCCATGCTCCATGCCCTCCCTCCGCGACGTCATCAC 1172  
 Db 121 PheArgAspLeuAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

QY 1173 GCGGACTTTGACAAATCAGCAGGAGCTGGAGATCTTCTCAACACATTCGCTACCCGAGC 1232  
 Db 141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

QY 1233 TCCTACGCCAACCGCTCTCTCCGCGTCAATCGGTAGAGACGACGAGACCCCTCATCGAG 1292  
 Db 161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

QY 1293 GAGCTCAATCCCGGCGAGCGCTTGGAGCTGTAGGCGCGGGGCGACAGGGGTGTGTGACC 1352  
 Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValThr 200

QY 1353 GACTTCGACGAGACGGGATGCTGGACCTCACTTGTCCATGGAGATGCTGCTCAG 1412

Db 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220  
 QY 1413 CCGCTGTCGCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGGAGTGTGCCA 1472  
 Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnTrpLeuArgValValPro 240  
 QY 1473 CGCACCCCGGTTGGGGCTTTGGCCAGGGGAGCTAAGGTGCTGCTCTACACCAAGAGAGT 1532  
 Db 241 ArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260  
 QY 1533 GGGCCCACTGAGGATCATCGACGGGGCTCAGGTACCTGTGTGAGATGAGAGCCCGTG 1592  
 Db 261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280  
 QY 1593 GCACATTTGGCCCTGGGAGGATGAAGCAGCAGTGTGGAGGTGACGTGCGCCAGATGCG 1652  
 Db 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300  
 QY 1653 AAGATGCTGAGCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATCTCTAC 1712  
 Db 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320  
 QY 1713 CCGCGGATGAGGACACACTTCAGGACCCCGCCACTGGAGTGTGGCCCAAGATTCTCC 1772  
 Db 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGlyPheSer 340  
 QY 1773 CAGCAGAAATGGCCATTCATGAGCAGCACCAATGAATGATCCAGTTCCTCCATTCGTGTC 1832  
 Db 341 GlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCys 360  
 QY 1833 CCTCGACACAAAGCCGCTATGTGTCAACACCTATGGAAGCTACAGGTGCGGACCAACAAG 1892  
 Db 361 ProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLys 380  
 QY 1893 AAGTGCAGTCCGGGCTACGAGCCCAACAGAGATGGCAGCCTGCGTGGGTGGTGAGC 1952  
 Db 381 LysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly----- 397  
 QY 1953 CCTGTGTGAAGATAGTGACACCAACAGTTGGGAAGACCTTGGTCCC 2000  
 Db 398 -----ThrLeuGlyGlnSerProGlyPro 405

RESULT 8  
 Q8R3V8 PRELIMINARY; PRT; 435 AA.  
 AC Q8R3V8;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Similar to cartilage acidic protein 1.  
 GN CRTAC1 OR 2810454P21RK.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC024472; AAH24472.1; -.  
 DR MGD; MGI:1920082; Crtac1.  
 DR GO; GO:0008305; C:integrin complex; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR InterPro; IPR001881; EGF CA.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR PROSITE; PS01187; EGF CA; 1.  
 SQ SEQUENCE 435 AA; 47315 MW; 01A18B8556A2F22B CRC64;

Alignment Scores:

```
Pred. No.: 1,09e-118 Length: 435
Score: 2035.50 Matches: 380
Percent Similarity: 94.94% Conservative: 14
Best Local Similarity: 91.57% Mismatches: 14
Query Match: 50.18% Indels: 7
DB: 11 Gaps: 1

US-09-914-958B-35 (1-2177) x Q8R3V8 (1-435)

QY 753 ATGGACCTCGAGCGCAGTACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGCTGCT 812
Db 1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20
QY 813 GAGGTGGGTGACGAATATACAGGGGGCGAGCGCTCAGCGTGGGCCCATCTCAGC 872
Db 21 GluAlaGlyValSerLysTyrThrAlaGlyArgGlyValSerValGlyProIleLeuSer 40
QY 873 AGCAGTGCCTCGGATATCTTCGACAAATGAGATGGCCCTAACTCTCTTTCCACAAC 932
Db 41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60
QY 933 CGGGCGCATGGCACCTTTGTGACGCTCGCGCAGCTGCTGGTGTGGACGACCCCCACACG 992
Db 61 GlnGlyAsnGlyThrPheValAspThrAlaAlaSerAlaGlyValAspAspProHisGln 80
QY 993 CATGGCGAGGTGTGCCCTCGCTGACTTCAACCGTGTGCGAAAGTGGACATCGTCTAT 1052
Db 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100
QY 1053 GCACTAGGAAGTGGCCCCCAGCGCTCTATCTGCAATGAGACACCCATGGAGGTCCGCG 1112
Db 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerAlaHisGlyLysValArg 120
QY 1113 TTCCGGGACATCGCTCACCCCAAGTTCTCATGCCCCCTCCGTCGCGCAGGTATCATCC 1172
Db 121 PheArgAspIleAlaSerProLysPheSerThrProSerProValArgThrValIleAla 140
QY 1173 GCGGCTTTGACAAATGACAGAGCTGGAGATCTTCTCAACACATCGCTTACCGCAGC 1232
Db 141 AlaAspPheAspAsnAspGlnGluLeuGluValPhePheAsnAsnIleAlaTyrArgSer 160
QY 1233 TCCTCAGCAACCGCTCTCCGCGTCATCCGTCAGAGACGACGACCCCTCATCGAG 1292
Db 161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180
QY 1293 GAGCTCAATCCCGGACGCTTTGGAGCTTGAGCGCGGCGACAGGGGTGTGTGACG 1352
Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200
QY 1353 GACTTCAGCGAGACGGGATGCTGACCTCATCTCTCCATGAGAGTCCATGGCTCAG 1412
Db 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220
QY 1413 CGCTCTCGCTCTCCCGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTGGTGCCA 1472
Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheSerAsnAsnTrpLeuArgValPro 240
QY 1473 CGCACCCGTTTGGGCTTTTGGCGAGCTTAAGTCTGCTCTACACCAAGAGAGT 1532
Db 241 ArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysSer 260
QY 1533 GGGGCCCATCTGAGGATCATCGAGCGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTG 1592
Db 261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280
QY 1593 GCACACTTTGGCTGGGAAAGTGAAGCCAGCAGTGTGGAGGTGAGCTGGCCAGATGGC 1652
Db 281 AlaHisPheGlyLeuGlyArgAspGluAlaSerSerValGluValThrTrpProAspGly 300
QY 1653 ARGATGTGAGCGGAACTGTGGCGAGCGGGAGATGAATCACTCAGTGTGGAGATCCTTAC 1712
Db 301 LysMetValSerArgSerValAlaSerGluGluMetAsnSerValLeuGluIleLeuTyr 320

1713 CCCGGGATGAGGACACACTTTCAGAACCCAGCCCGGATGGAGTGGCCAGGATTCTCC 1772
Db 321 ProGlnAspGluAspLysLeuGlnAsnThrAlaProLeuGluCysGlyGlnGlyPheSer 340
QY 1773 CAGCAGGAAATGCGCATTCATGATGACCAATGATGATGATGATGATGATGATGATG 1832
Db 341 GlnGlnAspAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCys 360
QY 1833 CCTCGAGCAACCGGATGTGTCAACACCTATGGAAGCTACAGGTGCGGACCAACCAAG 1892
Db 361 ProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLys 380
QY 1893 AAGTCACTCGCGGCTACGAGCCCAACGAGATGACAGCAGCTCGCTGCTG 1940
Db 381 ArgCysAsnArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGlnValAla 400
QY 1941 -----GGCTGTGGAGCCCTGTGTGAAGATAGTGACACCA 1976
Db 401 PheLeuGlyGlyTyrSerSerAlaAlaPheArgLeuSerGluPro 415

RESULT 9
Q9NW46 PRELIMINARY; PRT; 418 AA.
AC Q9NW46;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ10320.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira K., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001182; BAA91540.1; -.
DR GO; GO:0008305; C:intrinsic complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 3.
KW Hypothetical protein.
SQ SEQUENCE 418 AA; 4861 MW; B85F301A96DD545F CRC64;

Alignment Scores:
Pred. No.: 1,37e-114 Length: 418
Score: 1967.50 Matches: 385
Percent Similarity: 89.81% Conservative: 3
Best Local Similarity: 89.12% Mismatches: 10
Query Match: 48.51% Indels: 34
DB: 4 Gaps: 2

US-09-914-958B-35 (1-2177) x Q9NW46 (1-418)

QY 753 ATGGACCTCGAGCGCAGTACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGCTGCT 812
Db 1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20
QY 813 GAGGTGGGTGACGAATATACAGGGGGCGAGCGCTCAGCGTGGGCCCATCTCAGC 872
Db 21 GluAlaGlyValSerLysTyrThrAlaGlyArgGlyValSerValGlyProIleLeuSer 40
QY 873 AGCAGTGCCTCGGATATCTTCGACAAATGAGATGGCCCTAACTCTCTTTCCACAAC 932
Db 41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60
```

```
QY 933 CGGGCGATGGACCTTTGTGGACGCTGGGCGCAGTCTGTGTGGACGACCCCGACGAG 992
Db 61 ArgGlyAspGlyThrPheValAspAlaAlaSerAlaGlyValAspProHisGln 80
QY 993 CATGGCGAGGTGTGGCCCTGGCTGACTCAACCGCTGATGGCAAGTGGACATGCTCTAT 1052
Db 81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100
QY 1053 GGCACCTGAATGGCCCCCGCCATCGCCTCTATCTGCAAAATGAGCACCCTATGGGAAGTCCGC 1112
Db 101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120
QY 1113 TTCGGGACATCGCTCACCAAGTTCCTCATGCTCCCTCCCTGTCGCGACGCTATACCC 1172
Db 121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140
QY 1173 GCCGACTTTGCAATGACCGAGCTGGAGATCTTCTTCAACAATGCTTACCGCAGC 1232
Db 141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160
QY 1233 TCCTCAGCCACCGCTCTTCGGCGTGTATCCGTAGAGCAGGAGACCCCTCATCGAG 1292
Db 161 SerSerAlaAsnArgLeuPheArgValIleArgGluHisGlyAspProLeuIleGlu 180
QY 1293 GAGCTCAATCCCGCGCAGCTTGGAGCTGAGCGCTGAGCGCGGCGCACAGGGGTGTGTGACC 1352
Db 181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyValValThr 200
QY 1353 GACTTCGACGAGAGCGGATGTGGACCTCATCTTGTCCCATGAGAGTCCATGCTCAG 1412
Db 201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220
QY 1413 CCGCTGTCCTCTTCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGTGTCGA 1472
Db 221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240
QY 1473 CGCACCCGGTTTGGGGCTTGTGGCGGAGCTAGCTAGCTGTGCTGTCTACACAGAAAGT 1532
Db 241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValLeuTyrThrLysLysSer 260
QY 1533 GGGGCCCCACCTCAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTG 1592
Db 261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280
QY 1593 GCACACTTGGCTGGGAAGATGAAGCCAGCAGTGTGGAGTGAGCTGGCCAGATGGC 1652
Db 281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300
QY 1653 AAGATGCTGAGCGGAACGTGGCCAGCGGGAGATGAACCTCAGTGTGGAGATCCTCTAC 1712
Db 301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320
QY 1713 CCGCGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGATTCCTC 1772
Db 321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu----- 334
QY 1773 CAGCAGGAAATGGCCATTGGCATGGACACCAATGAATGCATCCAGTTCCTCATTCGTGTGC 1832
Db 335 -----ThrpProMetAsnAlaSerSerHisSerCysAl 346
QY 1833 CCTCGAGACAAGCCGTATGTGCACACCTATGGAGCTACAGGTGCGGACCAACAAG 1892
Db 346 aleuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366
QY 1893 AAGTGCAGTGGGGCTTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGCTGGTGACC 1952
Db 366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaIleTrpGly----- 383
QY 1953 CCTGTGTTGAAGATAGTGCACCAAGAATTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
Db 384 -----LeuSerAlaSer 387
```

```
QY 2013 TCACCTGCCTTGAATCACCGCTGGGAATACCTGTT 2046
Db 387 rHisArgAlaProAlaProProProProProLeu 398
RESULT 10
Q90X63 PRELIMINARY; PRT; 193 AA.
AC Q90X63;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Kawasaki H., Housman D.E., Graybiel A.M.;
RT "Characterization of region-specific genes from rat brain.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U78304; AAF21221.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR001881; EGF_Ca.
DR PROSITE; PS01187; EGF_CA; 1.
SQ SEQUENCE 193 AA; 21572 MW; B94898BFE5EE045A CRC64;

Alignment Scores:
Pred. No.: 1,32e-48 Length: 193
Score: 910.00 Matches: 165
Percent Similarity: 96.67% Conservative: 9
Best Local Similarity: 91.67% Mismatches: 6
Query Match: 22.44% Indels: 0
DB: 11 Gaps: 0

US-09-914-958b-35 (1-2177) x Q90X63 (1-193)
QY 1404 ATGGCTCAGCGCTGTCGCTTCGCGGCAATCAGGGCTTCAACAACACTGGCTCGGA 1463
Db 1 MetAspGlnProLeuSerValPheArgGlyAsnGlnGlyPheSerAsnTrpLeuArg 20
QY 1464 GTGGTCCACGACCCCGGTTTGGGGCTTTCGAGGGAGCTAAGGTGCTGTCTACACC 1523
Db 21 ValValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThr 40
QY 1524 AAGAAGTGGGGCCCACTCAGGATCATCGACGGGGCTCAGGCTACCTGTGTGAGATG 1583
Db 41 LysLysSerGlyAlaHisLeuArgIleIleAspGlySerGlyTyrLeuCysGluMet 60
QY 1584 GAGCCGCTGGCACACTTTCGCTGGGAGAGTGAAGCCAGCAGTGTGGAGGTGACGTGG 1643
Db 61 GluProValAlaHisPheGlyLeuGlyArgAspGluAlaSerSerValGluValThrTrp 80
QY 1644 CCAGATGGCAAGATGGTGGACCGGACGTCGGCGAGAGTGAACCTCAGTCTGGAG 1703
Db 81 ProAspGlyLysMetLeuSerArgSerValAlaAsnArgGluMetAsnSerValLeuGlu 100
QY 1704 ATCTCTACCCCGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGCCAA 1763
Db 101 IleLeuTyrProArgAspGluAspLysLeuGlnAsnThrAlaProLeuGluCysGlyGln 120
QY 1764 GGATCTTCCCGACGAGAAATGGCCATTGTCATGGACACCAATGAATGATCCAGTTCCTCA 1823
Db 121 GlyPheSerGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPhePro 140
QY 1824 TTCGTGTCCCTCGAGACAAAGCCCGCTATGTGTGTCAACACTATGGAAGCTACAGGTGCCG 1883
Db 141 PheValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArg 160
QY 1884 ACCAACAAGATGTGAGTGGGGCTTACGAGCCCAACGAGGATGGCACAGCTGCGTGGGC 1943
Db 161 ThrAsnLysArgCysSerArgGlyTyrGluProAsnGluAspGlyThrSerLeuArgGly 180
```







709 AspAspSerAlaValArgLysProIleGluArgThrProAspGlyThrProValLysLeu 728  
 QY 780 CGGGGCTATCTGCGCTCAGAGATGTGCTCGAGGTGGGGTCCAGCAATATACAGG 839  
 Db 729 ProGlyProLeuHisPheGlnProAlaLeuSerGluValLeuValSerAsnGlyLysGly 748  
 QY 840 -----GGCCGAGGCGTC 851  
 Db 749 AlaGlyArgArgGlnSerLeuGlyAspSerAlaGlnSerArgSerThrGlyMetGlyVal 768  
 QY 852 AGCTGGGCCCCATCTCAGCAGCAGTCCCTCGGATATCTTCGGCAATGAGAATGGG 911  
 Db 769 MetIleAlaAspMetAspGlnGlnProGlyAsnGluIlePheValAlaAsnAspGlnArg 788  
 QY 912 CTTAACTTCTTTCACACCGGGCGATGGCACC-----TTTGTGGACGCT 959  
 Db 789 AlaAsnHisPheTrpPheGluAlaGlyAspAlaAsnGluGluGlnGluTrpArgAsnGln 808  
 QY 960 GCGGCCAGTGTGTGTG-----GACGACCCGCCACCATGGCGAGGTGTC 1007  
 Db 809 AlaAlaIleArgGlyValAlaValGlyProGlyGlyAspLysAsnAlaCysMetGlyIle 828  
 QY 1008 GCCCTGGCTGACTTCAACCGTATGCGCAAGTGGACATCTCTATGGCACTTGGAAATGCG 1067  
 Db 829 AlaMetAlaAspPheAsnArgAspGlnLysLeuAspLeuHisValSerAsnPheTrpAsp 848  
 QY 1068 CCCACCGCTCTATCTGCAATGAGCACCCATGGGAAGGTCCGCTTCGGGACATCGCC 1127  
 Db 849 GluTrpAlaAsnHisPheCysGlnThrSerAlaGlyMetPheAlaAspArgAlaValAla 868  
 QY 1128 TCACCCAGTTCCTCATCGCTCCCTGTCGCCACG-----GTCTATCCCGCC 1175  
 Db 869 PheGlyIleAspGlnLeuThrSer-----ArgThrThrGlyPheGlyValGlnGlyLeu 886  
 QY 1176 GACTTTGCAATGACAGGAGTGGAGATCTTCTTCAACAAATTCCTACCGCAGCTCC 1235  
 Db 887 AspTyAspAsnAsnGlyTrpIleAspLeuLeuValAlaAsnGlyHisValGluAspLeu 906  
 QY 1236 TCAGCCACCGCTCTTCGCGCTCATCCGTAGAGACACGGAGACCCCTCATCGAGGAG 1295  
 Db 907 Thr-----AspArgGlyThrProPheArgMetLys 916  
 QY 1296 -----CTCAATCCCGGCGAGCGCTGGAG-----CTGAG 1325  
 Db 917 ThrGlnLeuValAsnGlnGlyLysArgPheGluLeuAlaLysMetAspProAsp 936  
 QY 1326 GCGCGG-----GGCACAGGGGGTGTGTGACCGACTTCGACGGAGAC 1367  
 Db 937 GlyGlnTrpGluArgGluArgLeuGlyArgGlyValAlaThrCysAspPheAspArgAsp 956  
 QY 1368 GGGATGCTGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCGCTGTCGCTTC 1427  
 Db 957 GlyArgLeuAspAlaValThrTyrAlaAspSerAlaAlaGlu-----LeuLeu 973  
 QY 1428 CGGGGCAATCAGGGCTTCAACAACTGCTCGAGTG---GTGCCACCGCCCGTTT 1484  
 Db 974 ArgAsnGluThrGluAlaValGlyAsnTyrIleGlnLeuArgLeuIleGlyThrArgSer 993  
 QY 1485 GGGGCTTTGCCAGGGAGCTAAGTGTGCTCTACACCAAGAAGAGTGGGGCCCACTG 1544  
 Db 994 GluArgAspAlaIleGlyAlaHisIleGluLeu---ThrAlaGlyAspGlnValTrpThr 1012  
 QY 1545 AGGATCATCAGCGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTGGGCACACTTGGC 1604  
 Db 1013 SerValValThrAlaGlyAspGlyTyrAlaCysArgAsnGluAlaIleLeuHisPheGly 1032  
 QY 1605 CTGGGAGAGTGAAGCAGC-----AGTGTGGAGTGCCTGGCCAGATGCAAGAT 1658  
 Db 1033 LeuGly---AspValAlaThrLeuAspSerIleAlaValGlnTrpProSerGlyGluVal 1051  
 RESULT 13  
 Q7UP15 PRELIMINARY; PRT; 1028 AA.  
 ID Q7UP15

Q7UP15;  
 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Probable acid secreted protein ASPIC.  
 GN RB6914.  
 OS Rhodopirellula baltica.  
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 OC Planctomycetaceae; Pirellula.  
 OX NCBI\_TaxID=117;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=1;  
 RC MEDLINE=22735913; PubMed=12835416;  
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
 RA Schlesner H., Amann R., Reinhardt R.;  
 RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
 RT strain 1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
 DR EMBL; BX294145; CAD75077.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 1028 AA; 114349 MW; C06C31757BFFDF8B CRC64;  
 Alignment Scores:  
 Pred. No.: 7,77e-08 Length: 1028  
 Score: 260.50 Matches: 139  
 Percent Similarity: 40.15% Conservative: 79  
 Best Local Similarity: 25.62% Mismatches: 223  
 Query Match: 6.42% Indels: 103  
 DB: 16 Gaps: 28  
 US-09-914-958b-35 (1-2177) x Q7UP15 (1-1028)  
 QY 261 TATGACAGTAATCCACCACCGCTCACTATGCTGGCAGTTACTGATGTGGACCATGAT 320  
 Db 532 TyrGluSer-----LeuGlyGlyValGlyValIleAspTyrAspAlaAsp 547  
 QY 321 GGGGACTTTGAGATGCTGCTGGGGGTACAATGGACCCCAACCTGTTCTGAAGTATGAC 380  
 Db 548 GlyAsnProAspLeuTyrPheAlaGlnGlySerAlaAspProGluLeuArgGlyThr 567  
 QY 381 CGGGCCAGAGAGCGGTGTGTGAAC-----ATCGCGTGTGATGAGCGCAGC 425  
 Db 568 ArgProAsnGlnLeuPheArgAsnLeuGlyLysArgPheValAlaValGluGlnSer 587  
 QY 426 TCACCTACTACGCGTGGGACCGCGAGGGAACGCCATTGGGTTCACAGCTGGCAG 485  
 Db 588 Gln-----CysAspGluPheGlyTyrSerGlnGlyIleSerValGlyAsp 602  
 QY 486 ATCGACGGGACGCGCGGAGGAGATCTACTTCTCTCAAC-----ACCAATAATGCTTC 539  
 Db 603 ValAsnGlnAspGlyTrpProAspLeuLeuValAlaAsnIleGlyLeuAsnArgLeuPhe 622  
 QY 540 TCGGGGTGGCCACGTCACACCGCAAGTGTTCAGATTTCGCGCAATACCGGTGGGAAGAC 599  
 Db 623 IleAsnCysAlaAspGlyThr-----PheArg----- 631  
 QY 600 ATCCTGAGCGATGAGTCAACGTCGGCCGCTGTGTGGCCAGCTCTTTGCGGACGCTCT 659  
 Db 632 -----AspAlaThrAspGlnMetHisGlyGlnAspGluArgPheThrSer---Ser 647  
 QY 660 GTGGCTGTGTGGACAGAAAGGCTCTGGACGCTACTCTATCTATCTACATTCGAATTAC--- 716  
 Db 648 MetAlaIleAlaAspIleAspGlyAspAlaLeuProAspLeuPheGluValAsnTyrIle 667  
 QY 717 -----GCTACGGTAATGTGGCCCTGTGCTCCTC--- 746  
 Db 668 GlnLeuGluGlyAlaPheAspProGlnPheAspSerGlnGlyArgGluLeuGln 687  
 QY 747 -----ATTGNAATGGACCTGAGCCAGTGACCTCTCCGGGGCAATCTCTGGCGTCA 800  
 Db 688 GlyProLeuSerGlnLeuProGluSerAspHisTrpTyrArgGlyAspGlyArgGlyAsn 707





```
QY 496 ACGCCGGGAGAGATCTACTTCTCAACACCAATAATGCCTTCTCGGGGTGGCCAGT 555
Db   ::::: |||
QY 115 SerValGlnProSerSerGlySerValProThrThrSerAlaThrSer----- 131
Db   ::::: |||
QY 556 ACACCGCAAGTTTCAAGTTCCGCAATAACCGGTGGGAAGACATCTCTGAGCGATGAGG 615
Db   ::::: |||
QY 132 ValGlnSerSerSerSerAlaProThrThrSerAlaThrSer----- 147
Db   ::::: |||
QY 616 TCACGTGGCCGTGGTGGCCAGCCTTTTGGCGGAGCTCTGTGGCCCTGTGTGGACA 675
Db   ::::: |||
QY 148 -----ValGlnProSerSer----- 152
Db   ::::: |||
QY 676 GAAAGGCTCTGGACGTACTCTATCTACANTGCAATTAGCCTACGGTAATGTGGCC 735
Db   ::::: |||
QY 153 -----SerSerSerProProIleSerSerThrVal-----Ser 163
Db   ::::: |||
QY 736 CTGATGCCCTCATTTGAAATGGACCTCGAGGCCAGTGACCTCTCCGGGGCATTCGTGGCGC 795
Db   ::::: |||
QY 164 ValGlnProSer----- 167
Db   ::::: |||
QY 796 TCAGAGATGTGGCTGTGAGCTGGGTCTAGCAAAATATACAGGGGGCCGAGCGTCAAGC 855
Db   ::::: |||
QY 168 -----SerSerSerAlaProThrThrSerAlaThrSer 179
Db   ::::: |||
QY 856 TGGGCCCATCTCTCAGCAGCAGTCCTCGGATATCTTCT-----GGCACA 900
Db   ::::: |||
QY 180 ValGlnProSerSerSerProProIleSerSerThrValSerValGlnThrSer 199
Db   ::::: |||
QY 901 ATGAGAATGGCCCTAACTCTCTTTCCACAAACCGGGCGATGGCACCTTTGTGGACGCTG 960
Db   ::::: |||
QY 200 SerSerSerSerValProThrThrSerThrThrSerValGlnProSerSerSerSer 219
Db   ::::: |||
QY 961 CGCCGAGTGTGGTGGAGCAGCCCCCACCAGCAGTGGCGAGGTGTGCGCCTGGCTGACT 1020
Db   ::::: |||
QY 220 ValProThrThrSerAlaThrSerValArgSerSerSerSerSerSerThrProIlePro 239
Db   ::::: |||
QY 1021 TCAACCGTGTAGCAAGTGGACATCGTCTCTATGGCAACTGGATGGCCCCCAGCGCTCT 1080
Db   ::::: |||
QY 240 SerThrThrSerValGlnProSerSerSerSer-----AlaProThrThrSer 256
Db   ::::: |||
QY 1081 ATCTGAAATGACACCCATCGGAAGTCCGCTTCGGGGACATCGCTCACCCCAAGTTCT 1140
Db   ::::: |||
QY 257 AlaThrSerValGlnPro-----SerSerSerSerSerThrProIleProSerThr 273
Db   ::::: |||
QY 1141 CATGCGCTCCCTGTCCGACCGTCAATACCCCGCACTTTTGACAATGACAGAGCTGG 1200
Db   ::::: |||
QY 274 ThrSerValGlnProSerSerSerSerAlaProThrThrSerAlaThrSerValGln 293
Db   ::::: |||
QY 1201 AGATCTTCTCAACAACANTGCCCT-----ACCGCAGCTCTCAGCCAAACCGCTCT 1251
Db   ::::: |||
QY 294 ProSerSerSerSerProProIleSerSerThrIleSerValGlnProSerSerSer 313
Db   ::::: |||
QY 1252 TCGCGTCACTCCGTAGACGACGAGAC-----CCCTCATCGAGGAGCTCAATCCCG 1305
Db   ::::: |||
QY 314 SerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySerAlaPro 333
Db   ::::: |||
QY 1306 GCGAGCCTTGGAGCCTGAGGCCGGGCACAGGGGTGTGTGACCGACTTCGACGGAG 1365
Db   ::::: |||
QY 334 ThrThr-----SerAlaThrSer 339
Db   ::::: |||
QY 1366 ACGGATGTGGACCTCATCTGTCCATGGAGAGTCCATGGCTAGCGCTGTCCGTCT 1425
Db   ::::: |||
QY 340 ValGlnProSerSerSerProProIleSerSerThrIleSerValGlnProSer 359
Db   ::::: |||
QY 1426 TCGGGGCAATCAGGCTTCAACAACACTGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1485
Db   ::::: |||
QY 360 SerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySer 379
Db   ::::: |||
QY 1486 GGGCCTTCCAGGGGAGTAAGTCTGTCTACACCAAGAGAGTGGGGCCCACTGA 1545
Db   ::::: |||
QY 380 AlaPro-----ThrThrSerAlaThrSerValGlnProSerSer 392
Db   ::::: |||
```

```
QY 1546 GGATCATCGAG-----GGGGTCTAGCTACCTGTGTGTGATGGAGCCGCTGGCACACTTTG 1602
Db   ::::: |||
QY 393 SerSerSerValProThrThrSerAlaThrSerValArgSerSerSer----- 408
Db   ::::: |||
QY 1603 GCCTGGGGAGGATGAACCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGA 1662
Db   ::::: |||
QY 408 ----- 408
Db   ::::: |||
QY 1663 GCCGGAACGTGGCCAGCGGGGAGATGAACCTCAGTGTGGAGATCTCTTACCCCGGGATG 1722
Db   ::::: |||
QY 409 SerSerSerThrProIleProThrThrThrSerValGlnProSerSerSerSerVal 428
Db   ::::: |||
QY 1723 AGGACACATTCAGGACCCAGCCCACTGGAGTGTGGCCAAAGGATTTCCACAGCAGGAAA 1782
Db   ::::: |||
QY 429 ProThrThrSerAlaThr-----SerValGlnThrSerSerSerSerThr 444
Db   ::::: |||
QY 1783 ATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCTCCATTCGTGT----- 1830
Db   ::::: |||
QY 445 ProIleProSerThrThrThrSerValGlnProSerSerSerSerSerAlaProThrThrSer 464
Db   ::::: |||
QY 1831 -----GCCCTCGAGACACAGCCCGTATGTGTCAACACTATGGAAGCTACA---GGTGCC 1881
Db   ::::: |||
QY 465 AlaThrSerValGlnProSerSerSerSerSerProIleSerSerThrIleSerVal 484
Db   ::::: |||
QY 1882 GGACCAACAAGAAGTGCAGTGGGGCTACGAGCCCAACAGAGATGGCAGCAGCTGCGTGG 1941
Db   ::::: |||
QY 485 GlnProSerSerSerSerSerProThrThrThrSerThrThrSerValGlnProSerSer 504
Db   ::::: |||
QY 1942 GCTGGTGGAGCC 1953
Db   ::::: |||
QY 505 SerGlySerAla 508
```

Search completed: August 3, 2004, 19:47:58

Job time : 142 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 17:08:04 ; Search time 5087 Seconds  
(without alignments)

12779.628 Million cell updates/sec

Title: US-09-914-958B-35

Perfect score: 2177

Sequence: 1 cggaggctcgagccagcc.....tcctcttgaaaaa 2177

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estm.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estio.\*

8: em\_estc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_est3.\*

12: gb\_est4.\*

13: gb\_est5.\*

14: gb\_est6.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_pbg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1544.2	70.9	2089	11 AK049801	AK049801 Mus muscu
2	1539.4	70.7	2019	11 AK083221	AK083221 Mus muscu
3	1483.6	68.1	2483	11 AK032328	AK032328 Mus muscu
4	1438	66.1	2147	11 BC042687	BC042687 Mus muscu

5	890.8	40.9	1022	13	BX460454
6	798	36.7	841	12	BG708845
7	790.8	36.3	904	13	BX451456
8	773.8	35.5	781	12	B1550776
9	773.6	35.5	1101	13	BQ067249
10	715.8	32.9	770	12	B1758703
11	708.4	32.5	804	12	B1458314
12	692.2	31.8	860	12	B1819378
13	682.6	31.4	926	10	BF527664
14	655	30.1	1063	12	B1596960
15	642.4	29.5	656	13	BQ183029
16	634.6	29.2	735	12	BG911024
17	628.8	28.9	636	10	BF727258
18	600.4	27.6	713	9	AI792073
19	590.4	27.1	740	12	BM948537
20	574.8	26.4	578	14	CD676152
21	574.4	26.4	731	13	BQ178121
22	570.6	26.2	580	12	B1549399
23	566	26.0	567	12	BM707643
24	556.8	25.6	608	9	AU123536
25	556.6	25.6	729	12	BG818593
26	548.4	25.2	551	12	BM699706
27	547.8	25.2	622	12	B1824952
28	541.4	24.9	749	12	BG976852
29	537.4	24.7	540	14	CB153273
30	537.4	24.7	679	10	BB429618
31	534.2	24.5	539	14	CB152752
32	532.4	24.5	566	12	BG721807
33	526.8	24.2	549	12	BM706915
34	525.2	24.1	622	12	BG716170
35	524.4	24.1	721	14	CK221925
36	505.4	23.2	754	14	CK221936
37	498.2	22.9	559	10	BF080090
38	487.6	22.4	1418	12	BM805494
39	486	22.3	598	13	BQ637467
40	477	21.9	511	9	AI791928
41	474.4	21.8	489	10	BF725127
42	470.6	21.6	554	10	BF080091
43	467.4	21.5	508	10	BF088585
44	462.2	21.2	661	13	BY741935
45	459	21.1	660	10	BB653351

#### ALIGNMENTS

#### RESULT 1

AK049801

LOCUS

DEFINITION

AK049801

2089 bp mRNA

linear

HTC 20-SEP-2003

Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length

enriched library, clone:C530050M20 product:ASPIC PRECURSOR homolog

[Homo sapiens], full insert sequence.

AK049801

GI:26340523

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

3  
REFERENCE  
AUTHORS  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Iihikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
MEDLINE  
PUBMED  
11076861  
4  
REFERENCE  
AUTHORS  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 695-690 (2001)  
5  
REFERENCE  
AUTHORS  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6  
REFERENCE  
AUTHORS  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
TITLE  
JOURNAL  
COMMENT  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/.  
FEATURES  
source  
Location/Qualifiers  
1. .2089  
/organism="Mus musculus"  
/mol\_type="mrna"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:C530050M20"  
/db\_xref="MGI:2417318"  
/db\_xref="taxon:10090"  
/clone="C530050M20"  
/tissue\_type="spinal cord"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
122. .2062  
/note="unnamed protein product; ASPIC PRECURSOR homolog  
[Homo sapiens] (SPTR|Q9NQ80, evidence: FASTA, 94.1%ID,  
91.6%length, match=1821)  
putative"  
/codon\_start=1  
/protein\_id="BAC33924.1"  
/db\_xref="GI:26340524"  
CDS

/translation="MAPSADPQMVRMALLPPLMLLELTGGSQRAEPMFTAVTNSVL  
PPVDNSPTQNLNGVAVTDVHDGDFEITWAGYDTPNLVLKYNRAQNLVNIADVERS  
SPYALDRDQNAIGTACDIDGGRBRIYFLNTNNAFSGVATYDKLKFERNNEWED  
ILSDVNVARGVASLIFAGRSVACVDRITGSGRSYIYANAYDVGDPDALIEMDPEASD  
LSRGLILRDVAAAGVSKYTAGRGVSVPGLISDVSAGNPNFLPHNCGT  
FVDTAAGAGVDDPHQGRGVALDFNRDKVDIYVGNWNGPHRLFLQMSAIGKVRPRD  
IASPKFSTPSPRTVIAADFNDQDELFVFNFIYRSSANLRFVIRREHNGPLLEE  
LNPGDALPEGRGTGVVTFDGDMLDLILSHGSMAPLSVFEFGNFGNNLURVV  
PRTFPGAPCAKVVLYTKSGAHLRIIDGSGYLCEMEPVAFHGLGRDEASSVEVTW  
PDGMVSRVASEEMNSVLEILLYPQDEDLONTAPLECGQFSGDQNDHMDTNCIQ  
PFVFCPRDKPVSEVNTYGSIRCTNRCNRGTFEPNEDGTACVAQVAFLOGYSAAARLS  
EPLSQASYLSLGLGLCLQIAL"  
2071. .2076  
/note="putative"  
2089  
/note="putative"  
polyA\_signal  
polyA\_site  
ORIGIN  
Query Match 70.9%; Score 1544.2; DB 11; Length 2089;  
Best Local Similarity 88.2%; Pred. No. 3.9e-287;  
Matches 1714; Conservative 0; Mismatches 223; Indels 6; Gaps 3;  
QY 5 AGGCTCGAGGCGAGCCGGGACCGGGCTGGAGCAAGCAGCGGGCGGCGCGGGCAG 64  
DB 1 AGGCTCGAGGCGAGCCGGGACCGGGCTGGAGCGCTGCAAGCGCGGCGGCGGCGAG 60  
QY 65 AGGCGCGAGCGAGCGCCGCGC-TTCCACCGCCCTAGGCGCGCGGGGCGGCGGCGAG 121  
DB 61 AGGCGAGCATCGAGCGCGCGCGCTCGACCGCCCGCGGCTGGAGCGGCGGCGAG 120  
QY 122 GATGGCTTCGAGCGCTGACCCCGGCGATGTCAGGATGTTTACCGTTCCTGCTGTGCT 178  
DB 121 GATGGCTTCGAGCGCTGACCCCGGCGATGTCAGGATGTTTGTCTGCTGCGCGCCCT 180  
QY 179 CTGGTTCGCGCATCATCTAGAGGGTCCAGCGGGCTGAACCCATGTTTCACTGCATCTAC 238  
DB 181 GTGGCTTCGCTCTCACTGGGGGATCCAGAGGGCTGAGCCCATGTTCACTGCAGTTCAC 240  
QY 239 CAACTCAGTTCCTGCTCTGACTATGACAGTAAATCCACCCAGCTCAACTATGTTGGC 298  
DB 241 CAACTCAGTTCCTGCTCTGACTATGACAGTAAATCCACCCAGCTCAACTATGTTGGC 300  
QY 299 AGTTACTGATGTGGACCATGATGGGACCTTTGAGATGCTGTCGGGGGTACAAATGGACC 358  
DB 301 AGTGACAGATGTGGACCATGATGGGACCTTTGAGATGCTGTCGGGGGTACAAATGGACC 360  
QY 359 CAACTGTTTCTCAAGTATGACCGGCGCCAGAGCGGCTGTTGAAATCGCGTGCATGA 418  
DB 361 CAACTGTTTCTCAAGTATGACCGGCGCCAGAGCGGCTGTTGAAATCGCGTGCATGA 420  
QY 419 GCGCAGCTCACCTTACTAGCGCTCGCGGACCGGCGAGGGGAAACCCATTGGGCTCACAGC 478  
DB 421 GCGCAGCTCACCTTACTAGCTCTGAGGACCGTCAAGGGGAAACCCATTGGGCTCACAGC 480  
QY 479 CTGCGCATCGACGGGGACCGCGGAGGAGATCTTCTTCTCTCAACACCAATATGCTT 538  
DB 481 CTGTGACATCGATGGGACCGCGGCTGAGGAGATCTTCTTCTCAACACCAATATGCTT 540  
QY 539 CTGCGGGGTGGCCACGTACACCGACAAGTGTTCAGTTCGCGCAATAACCGTGGGAAGA 598  
DB 541 CTGAGGTGGCCACATACACAGCAAGTGTTCAGTTCGCGCAATAACCGTGGGAAGA 600  
QY 599 CATCTCTGAGCGATGAGGTCAACGTCGCGCGCTGTCGGCAGCGCTCTTTTCCCGGACGCTC 658  
DB 601 CATCTCTGAGCGATGAGGTCAACGTCGCGCGCTGTCGGCAGCGCTCTTTTCCCGGACGCTC 660  
QY 659 TGTGGCTGTGTGACAGAAAGGCTCTGACGCGTACTCTATCTATCAATTGCGCAATTAGC 718  
DB 661 CGTGGCTGTGTGACAGAAAGGCTCTGACGCGTACTCTATCTATCAATTGCGCAATTAGC 720  
QY 719 CTACGCTAATGTGGCGCTGATGCCCTCATTTGAAATGAGCGCTGAGCGCAGTGCCTCTC 778  
DB 721 CTATGCTAATGTGGCGCTGATGCCCTCATTTGAAATGAGCGCTGAGCGCAGTGCCTCTC 780



QY 779 CCGGGGCAATTCGTGGGCTCAGAGATGTGGCTGTGAGGCTGGGCTCAGCAAAATATACAGG 838  
Db 781 CCGGGGGATCCTGGCACTCAGGGACGTGGCGGCTGAGGCTGGGCTCAGCAAGTATACAGC 840  
QY 839 GGGCCGAGGGCTCAGCGTGGGCCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGGA 898  
Db 841 GGGCCGGGTGTAGCGGTAGGCCCATCTCTCAGCAGCAGTGCCTCAGATATCTTCTGGA 900  
QY 899 CAATGAGAAATGGGCTAACTCTCTTTTCCACAACCGGGCGATGGCACCTTTGTGGACGC 958  
Db 901 CAACGAGAACGGGCCCAACTCTCTCTTCCACAACCAAGGCAACGTAACCTTTGTGGATAC 960  
QY 959 TGGCCGACGTGCTGTGTGACGACGCCACACAGCATGGCGAGGTGTGCGCTGTGATGA 1018  
Db 961 TGGGCCAGTGGCGGCTAGACGACCCCTCATCAGCATGGCGAGGTGTGCGCTGTGCGAGA 1020  
QY 1019 CTTCAACCGTGATGGCAAGTGGACATCGTCTATGGCACTGGAATGGCCCGCCACCGCCT 1078  
Db 1021 TTTCAACCGTGACGGCAAGTAGACATCGTCTATGGCACTGGAACGGCCACATCGCCT 1080  
QY 1079 CTATCTGCAATGAGCAACCCATGGGAAGTCCGCTTCCGGACATCGCCTCACCCAAAGTT 1138  
Db 1081 CTATCTACAGATGAGTCTCAGGGGAAGTCCGCTTCCGGACATTCGTCGCCAAGTT 1140  
QY 1139 CTCATGCTCCCTCCCTGTGCGCAGGTGATCATCCGCGACTTTGACAAATGACAGGAGCT 1198  
Db 1141 CTCACGCTCCCTCCCTGTGCGGACTGTCAITGTGCGGACTTTGACAAATGACAGGAAT 1200  
QY 1199 GGAGATCTTCTTCAACAACATTCGCTACCGAGCTCTCAGCCAAACCGCTCTTCGCGT 1258  
Db 1201 GGAAGTCTTCTTCAACAACATTCGCTACAGAGCTCTCAGCCAAATCGCTCTTCGCTGT 1260  
QY 1259 CATCCGTGAGAGACGAGAACCCCTCATCAGGAGCTCAATCCCGGCAACGCTTTGGA 1318  
Db 1261 CATCCGAGGAGACGAGGACCTCTCATCGAGGAGCTCAATCCTGTGACGCCCTAGA 1320  
QY 1319 GCCTGAGGCGGGGACAGGGGTGTGTGACGACCTTCAGCGAGACGGGATGCTGGA 1378  
Db 1321 GCCTGAGGCGGGGTACAGGGGCGGTAGTACCGACTTCGATGGTGTGATGGGATGCTGGA 1380  
QY 1379 CCTCATCTTCTCCATGAGAGTCCATGGCTCAGCGCTGTCTGCTCTTCGGGGCAATCA 1438  
Db 1381 TCTCATCTTGTACATGGAGTCCATGGCTCAGCCACTGTCTCTTCGGGGAATCA 1440  
QY 1439 GGGCTTCAACAACTGGCTGAGGTGTGACCGACCCCGTTTGGGCTTTTGGCCAG 1498  
Db 1441 GGGCTTCAACAACTGGTGTGCTGTGTGTACCTCGCACCCGTTTCGGGCTTTTGGCAG 1500  
QY 1499 GGGAGCTAAGTGTGTCTTACACCAAGAGAGTGGGCGCCACCTGAGGATCATCGACGG 1558  
Db 1501 GGGCGCAAGGTGTGTCTTACACCAAGAGAGTGGGCGCCACCTGAGGATCATGATGG 1560  
QY 1559 GGGCTCAGGCTACTGTGTGAGATGGAGCCCTGGCACACTTTGGCTGGGGAGGATGA 1618  
Db 1561 GGGTTCGGCTACTGTGTGAATGAGGCTGTGGACATTTTGGCTGGGACGGGATGA 1620  
QY 1619 AGCCACAGTGTGGAGTGAAGTGGCCAGATGGCAAGATGGTGAAGCGGAACTGGGCCAG 1678  
Db 1621 AGCCACAGTGTGGAGTGAAGTGGCCAGATGGCAAGATGGTGAAGTGTGGCCAG 1680  
QY 1679 GGGGGAGTGAATCAGTGTGAGATCTCTTACCCCGGGATGAGACACATTCAGGA 1738  
Db 1681 CGAGGAGATGAATCAGTGTGTGGATCTCTTACCCCGGGATGAGACAACTTCAGAA 1740  
QY 1739 CCAGCCCCACTGGAGTGTGGCAAGGATTTCTCCAGCAGGAAATGGCCATTCGATGA 1798  
Db 1741 CACAGCCCCACTAGTGTGGCCCAAGATTTCTCCAGCAGGAAATGGCCATTCGATGA 1800  
QY 1799 CACCAATGAATCAGTGTCCATTCGTTGTGCTTCGAGCAAGCCCGTATGTCTCAA 1858  
Db 1801 CACCAATGAATCAGTGTCCATTCGTTGTGCTTCGAGCAAGCCCGTATGTCTCAA 1860

QY 1859 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAGTGCAGTCCGGGCTACGAGCCCAA 1918  
Db 1861 CACCTATGGAAGCTACAGGTGCCGGACCAATAAAGATGCAATCGGCGCTATGAACCCAA 1920  
QY 1919 CGAGATGGGCACAGCTCCGCTGG 1941  
Db 1921 TGAAGCGGCACAGCCTGTGTGG 1943  
RESULT 2  
AK083221 2019 bp mRNA linear HTC 20-SEP-2003  
LOCUS Mus musculus adult male hippocampus cDNA, RIKEN full-length  
DEFINITION enriched library, clone:G630029D04 product:ASPC PRECURSOR homolog  
[Homo sapiens], full insert sequence.  
ACCESSION AK083221  
VERSION AK083221.1 GI:26101131  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes  
MEDLINE Genome Res. 10 (10), 1617-1630 (2000)  
PUBMED 20493374  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Iogawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
JOURNAL sequencing pipeline with 384 multicapillary sequencer  
MEDLINE Genome Res. 10 (11), 1757-1771 (2000)  
PUBMED 20530913  
REFERENCE 4  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
REFERENCE 5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
JOURNAL of 60, 770 full-length cDNAs  
MEDLINE Nature 420, 563-573 (2002)  
PUBMED 12000000  
REFERENCE 6  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

**TITLE** Direct Submission  
**JOURNAL** Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
**COMMENT** cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://phantom.gsc.riken.go.jp/

**FEATURES****source**

1..2019  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:C630029D04"  
/db\_xref="MGI:2417602"  
/db\_xref="taxon:10090"  
/clone="C630029D04"  
/sex="male"  
/tissue\_type="hippocampus"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
1..2019  
/note="ASPIC PRECURSOR homolog [Homo sapiens]  
(SPTR|Q9NQ80, evidence: FASTV, 94.1%ID, 91.6%length, match=1821)"

**misc\_feature****ORIGIN**

Query Match 70.7%; Score 1539.4; DB 11; Length 2019;  
Best Local Similarity 88.1%; Pred. No. 3.2e-286;  
Matches 1722; Conservative 0; Mismatches 226; Indels 7; Gaps 4;  
  
QY 1 CGGAGGCTCGAGGCGAGCCCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGG 60  
DB 29 CGGAGGCTCGAGGCGAGCCCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGG 88  
QY 61 GCAGAGCGCGAGCGCGCGCGGCTTCCACGCCCTAGCGCGCGCGCGGCG 117  
DB 89 GCAGAGCGCGAGCGCGCGCGGCTTCCACGCCCTAGCGCGCGCGCGGCG 148  
QY 118 GGAGGATGCTCGAGCGCTGACCCCGCATGTCCAGGATGTATCCGTTCTCTGCTG- 176  
DB 149 AGAGGATGCTCGAGCGCTGACCCCGCATGTCCAGGATGTATCCGTTCTCTGCTGCGCG 208  
QY 177 --CTCTGTTTCGCCCATCTAGCGGCTCCAGCGGCTGAAACCATGTTCACTGCGAG 234  
DB 209 CCCTGTGGTCTTCTGCTCTCACTGGGGATCCAGAGGCTGAGCGCATGTTCACTGCGAG 268  
QY 235 TCACCAACTCAGTCTTCCCTCTGACTATGACAGTAATCCACCCAGCTCACTATGCTG 294  
DB 269 TCACCAACTCAGTCTTCCCTCTGACTATGACAGTAATCCACCCAGCTCACTATGCTG 328  
QY 295 TGGCAGTTACTGATGTGGACCATGATGGGACATTTGAGATCGTGTGGCGGGGTACAATG 354  
DB 329 TGGCAGTTACTGATGTGGACCATGATGGGACATTTGAGATCGTGTGGCGGGGTACAATG 388  
QY 355 GACCCAACTGTTCTGAAGTATGACCGGCGCCAGAGCGGCTGTGGTGAACATCGCGGTG 414  
DB 389 GCCCCAACTGTTCTGAAGTATGACCGGCGCCAGAGCGGCTGTGGTGAATTTGCGGTG 448  
QY 415 ATAGCGCAGCTCACCTTACTAGCGCTGGGACCGCGGAGCGGCGGCGGCGGCGGCGG 474  
DB 449 ACAGCGCAGCTCACCTTACTAGCGCTGGGACCGCGGAGCGGCGGCGGCGGCGGCGG 508  
QY 475 CAGCCTGCGACATCGAGCGGCGGCGGCGGAGGAGATCTACTTCTCTCAACACCAATAATG 534  
DB 1589 GATGGGGGTTCCGGCTACCTGTGAAATGGAGCGCTGTGGACATTTTGGCCTGGGAGG 1648

DB 509 CAGCCTGTGACATCGATGGGACGCCCGTGGAGGATCTACTTTCTCAACACCAATAACG 568  
QY 535 CTTTCTCGGGGTGCCACGTACACCGACAAGTGTCTCAAGTTCGCGAATAACCGGTGG 594  
DB 569 CTTTCTCAGGTGTGGCCACATACACAGACAAGTGTCTCAAAATTCGGAATAACCGGTGG 628  
QY 595 AAGACATCTCAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTTGGCGAC 654  
DB 629 AAGACATCTCAGGTGATGACGTCAATGTGGCCCGTGGAGTGGCCAGCCTCTTTTGGCGAC 688  
QY 655 GCTCTGTGGCTGTGGACAGAAAGGCTCTGGAGCGTACTCTATCTATCTATCTGCAATT 714  
DB 689 GCTCGTGGCTGTGTAGACAGGACGGCTCTGGAGATCTCTATCTATCTATCTATCTGCA 748  
QY 715 AGCCCTACGTTAATGTGGCCCGTGTATGCTCTCATTTGAAATGACCTGAGGCGAGTGACC 774  
DB 749 ATGCTATGTTGATGTGGGCGCTGATGCCCTCATGAAATGACCTGAGGCGAGTGACC 808  
QY 775 TCTCCCGGGCAATCTCTGGCGCTCAGAGATGTGGCTGTGGCTGTGGGTGAGCAATAATA 834  
DB 809 TGTCCCGGGGATCTCTGGCACTCAGGAGCTGGCGCTGAGGCTGGGTGAGCAATAATA 868  
QY 835 CAGGGGCGGAGCGCTGAGCGTGGCGCCCATCTCTCAGCAGCAGTGGCTCGGATATCTTCT 894  
DB 869 CAGCGGCGCGGGCTGTCTAGCGTAGGCCCATCTCTCAGCAGCAGTGGCTCGGATATCTTCT 928  
QY 895 GCGCAATGAGAA-TGGGCTAACTTCTTTTCCACACCGGGCGATGGCACTTGTG 953  
DB 929 GTGCAACAGAAACCGGGCCCACTTCTTCTTCCACCAAGGCAACCGTACCTTGTG 988  
QY 954 GAGCGTGGCGGCAATCTCTGGTGTGGACGACCCCGACAGCATGGGCGAGGTGCGCCCTG 1013  
DB 989 GATCTGCGGCGAGTGGCGGCTAGACGACCTCATCAGCATGGCGAGGTGCGCCCTG 1048  
QY 1014 GTGACTTCAACCGTGTGAGCAAGTGGCAATCTCTATGCGCACTGGGAATGGCCCGCC 1073  
DB 1049 CAGAGATTTCAACCGTGTGAGCAAGTGGCAATCTCTATGCGCACTGGGAATGGCCCGCC 1108  
QY 1074 GCGCTCTATCTCAATGAGCAACCATGGAAAGTCCGCTTCCGGGACATCGCCTCACCC 1133  
DB 1109 CGCTCTATCTCAATGAGTGTCTACGGGAGGTCGGTTCGGGACATGCTTCGCGCC 1168  
QY 1134 AAGTTCTCCATGCTCCCTCTGCGACGCTCATCACCGCGCACTTTGACATGACCGAG 1193  
DB 1169 AAGTTCTCCACGCTCCCTCTGCGGACTGTCTATGCTGCGGACTTTGACATGACCGAG 1228  
QY 1194 GAGTGGAGATCTTCTTCAACCAATGCTTACCGGAGCTCTCTCAGCGACCGCCTCTTC 1253  
DB 1229 GAACCTGGAAGTCTTCTTCAACCAATGCTTACGAGGCTCTCTCAGCGCAATCGCCTCTTC 1288  
QY 1254 CGCGTCTATCGTAGAGACGAGACCCCTCTCATCGAGGAGCTCAATCCCGCGGACGCC 1313  
DB 1289 CGTGTCTATCGCAGGAGACCGGAGACCTCTCATCGAGGAGCTCAATCTCTGTGAGCGCC 1348  
QY 1314 TTGGAGCCTCAGGCGCGGGGCAAGGGGTGTGGTACCGGACTTCGAGGAGACGGGATG 1373  
DB 1349 CTAGAGCCTGAGCGCGGGGTACAGGGGCGGTAGTGACCGACTTCGATGATGGGATG 1408  
QY 1374 CTGGACCTCATCTTGTCCCATGAGAGATCGAGCTCAGCGGCTGCTCCGCTCTCCGGGG 1433  
DB 1409 CTGGATCTCATCTTGTGACATGAGAGTCCATGGCTCAGCCTGTCTGTCTTCCGGGGA 1468  
QY 1434 AATCAGGCTTCAACCAACACTGCGTGGAGTGGTGGTCCGACCGCGGTTTGGGGGCTTT 1493  
DB 1469 AATCAGGCTTCAACCAACACTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1528  
QY 1494 GCCAGGGGAGCTAAGTGTGTCTTACACCAAGAGTGGGCGCCCACTCTGAGGATCATC 1553  
DB 1529 GCCAGGGGCGCAAGGTGTGTCTTACACCAAGAGTGGGCGCGCACTTACCGATCAT 1588  
QY 1554 GACGGGGCTCAGGCTACCTGTGTGAGTGGCGGCTGGACACTTGGCCCTGGGGAAG 1613  
DB 1589 GATGGGGGTTCCGGCTACCTGTGAAATGGAGCGCTGTGGACATTTTGGCCTGGGAGG 1648

Qy	1614	GATGAAGCCACGACGTGTGGAGTGCAGCTGGCGCAGATGCCAAGATGGTCAGCGGAACGTG	1673
Db	1649	CATGAAGCCACGACGTGTGGAGTGCAGCTGGCGCAGATGCCAAGATGGTCAGCGGAAGTGTG	1708
Qy	1674	GCCAGCGGGAGATCAACTCAGTGTCTGAGATCTCTACCCCAGGATGAGGACAACACTTT	1733
Db	1709	GCCAGCGGGAGATGAATCTCGTGTTGAGATCTCTACCCCAGGATGAGGACAACACTTT	1768
Qy	1734	CAGGACCCAGCCCCACTGAGTGTGGCCAAAGATTCTCCACAGCAGGAAAAATGGCCATTGC	1793
Db	1769	CAGAACACAGCCCCACTAGAGTGTGGCCAAAGATTCTCCACAGCAGGAAATGGCCATTGC	1828
Qy	1794	ATGACACCATAATGATCATCCAGTGTCCATTTCGTTGTCCTTCGACAGAACCCCGTATGT	1853
Db	1829	ATGACACCATAATGATCATCCAGTGTCCATTTCGTTGTCCTTCGACAGAACCCCGTATGT	1888
Qy	1854	GTCAACACCTTATGGAAGCTACAGTGTCCGGCAACACAGAGTGCAGTTCGGGGCTACGAG	1913
Db	1889	GTCAACACCTTATGGAAGCTACAGTGTCCGGCAACACAGAGTGCAGTTCGGGGCTACGAG	1948
Qy	1914	CCCAACGAGGATGCACAGCGCTCGTGGGCTGGTG	1948
Db	1949	CCCAATGAAGACGGCACAGCCTGTGTGGGCTGGG	1983
RESULT_3	AK032328		
LOCUS	AK032328	Mus musculus adult male olfactory brain cDNA, RIKEN full-length	HTC 18-SEP-2003
DEFINITION		[Homo sapiens], clone:6430520G02 product:ASPIC PRECURSOR homolog	
ACCESSION	AK032328		
KEYWORDS		HTC; CAP trapper.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE	1	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shihata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE		High-efficiency full-length cDNA cloning	
JOURNAL		Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE		99279253	
PUBMED		10349636	
REFERENCE	2		
AUTHORS		Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shihata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL		Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE		20499374	
PUBMED		11042159	
REFERENCE	3		
AUTHORS		Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura,S., Hazama,M., Nishine.T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsushita,S., Kawai.J., Okazaki.Y., Muramatsu,M., Inoue,Y., Kira.A. and Hayashizaki,Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL		Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE		20530913	
PUBMED		11075861	
REFERENCE	4		
AUTHORS		The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
TITLE		Functional annotation of a full-length mouse cDNA collection	
JOURNAL		Nature 409, 685-690 (2001)	
REFERENCE	5		
AUTHORS		The FANTOM Consortium and the RIKEN Genome Exploration Research	

Db 4 GAGAGCCAGAGATGGCTCCGAGCGCTGACCGGGCATGGTCAGATGGCTTGTCTCTG 63  
QY 171 CTGTG---CTCTGTTTCTGCCATCACTAGAGGGTCCAGCGGGCTGAACCCATGTTTC 227  
Db 64 CTGCGCCCTGTGGCTTCTGCTCTCACTGGGGATCCAGAGGGCTGAGCCCATGTTTC 123  
QY 228 ACTGAGTACCAACTCACTGTTCTGCTCTGACTATGACAGTAATCCACCAGCTCAAC 287  
Db 124 ACTGAGTACCAACTCACTGTTCTGCTCTGACTATGACAGTAATCCACCAGCTCAAC 183  
QY 288 TATGTTGGGAGTTACTGATGTCACCATGATGGGACTTTGAGATCGTCTGTGGCGGG 347  
Db 184 TATGAGTGGGAGTGGAGATGTCAGATGATGAGGACTTTCGAGATCGTTGTGGCGGG 243  
QY 348 TACAAATGAGCCCAACTCTGTTCTGAAGTATGACCGGGCCCGGAGCGCTGGTGAACATC 407  
Db 244 TACACGGGCCCAACTCTGTTCTGAAGTATGACCGGGCCCGGAGCGCTGGTGAATAT 303  
QY 408 GCGGTGATGAGCGAGCTCACTGCTTACCGGCTGCGGGACCGGAGCGGAGCGCATT 467  
Db 304 GCGTGGAGAGCTCACTGCTTACCTGCTGAGGAGCGCTGAGGAGATCTACTTCTCAACACC 363  
QY 468 GGGTTCACAGCTGCGACATCAACCGGACCGGCGGAGGAGATCTACTTCTCAACACC 527  
Db 364 GGGTTCACAGCTGTCATCGATGGGAGCGGCGTGGAGGAGATCTACTTCTCAACACC 423  
QY 528 AATAATGCTTCTCGGGGTGCCAGTACACGCAAGTTGTTCAAGTTCCGCAATAC 587  
Db 424 AATAAGCCTTCTCAGGTGTGCCACATACACAGCAAGTTGTTCAAAATTCGGAATAC 483  
QY 588 CGGTGGAGAGATCTGAGCGATGAGTCAAGTGGCGGCTGGAGCGTACTCTATCTACATT 647  
Db 484 CGGTGGAGAGATCTGAGTATGACGTCATGTTGGCGGCTGGAGTGGCGGCTCTTT 543  
QY 648 GCGGAGCGTCTGTCGCTGTGACAGAAAGGCTCTGGAGCGTACTCTATCTACATT 707  
Db 544 GCGGAGCGTCTGTCGCTGTGACAGAAAGGCTCTGGAGCGTACTCTATCTACATA 603  
QY 708 GCCAATACCGCTAGCGTAAATGGGCGCTGATGCGCTCATTTGAATGACCGCTGAGGCC 767  
Db 604 GCCAATACCGCTAGCGTAAATGGGCGCTGATGCGCTCATTTGAATGACCGCTGAGGCC 663  
QY 768 AGTACCTCTCGGGGCAATCTGGCGCTCAGAGATGGCTGCTGAGCTGGGTCAGC 827  
Db 664 AGTACCTCTCGGGGCAATCTGGCGCTCAGAGATGGCTGCTGAGCTGGGTCAGC 723  
QY 828 AAATATACAGGGGCGGAGCGCTCAGCGTGGGCGGCTCATCTCAGCAGCAGTGCCTCGAT 887  
Db 724 AAGTACACAGGGGCGGAGCGCTCAGCGTGGGCGGCTCATCTCAGCAGCAGTGCCTCGAT 783  
QY 888 ATCTTCTCGCAATGAGAAATGGGCGCTAACTTCTCTTTTTCACACCGGGGCGATGCACC 947  
Db 784 ATCTTCTGTCACACGAGAAAGGCGGCGGCTTCTCTTTCACACCGGCAACGCTTACC 843  
QY 948 TTTGTGAGCGCTCGGCGGCTGCTGTTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1007  
Db 844 TTTGTGAGCGCTCGGCGGCTGCTGTTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 903  
QY 1008 GCGCTGGCTGACTTCAACCGTGTATGTCGAAAGTGGACATCGTCTATGGCAACTGGAATGGC 1067  
Db 904 GCGCTGGCGATTTCAACCGTGTATGTCGAAAGTGGACATCGTCTATGGCAACTGGAATGGC 963  
QY 1068 CCCACCGGCTCTATCTGAAATGAGACCCATGGGAGAGTTCGCTTCGGGACATCGCC 1127  
Db 964 CCACATCGGCTCTATCTACAGATGAGTGTCTACGGGAGGTTCCAGTTTCGGGACATTTGCT 1023  
QY 1128 TCACCAAGTTCTCCATCGCTCCCTGTCGCAAGTGTATACCGCGGCTTTGACAAT 1187  
Db 1024 TCGCCCAAGTTCTCCACGCGCTCCCTGTCGCAAGTGTATACCGCGGCTTTGACAAT 1083  
QY 1188 GAGCAGGAGTGGAGATCTTTTCAACAAATTCCTACCGGAGCTCTCTCAGGCAACCGC 1247

Db 1084 GACCAGGAATGGAAGTCTTCTTCAACAAATTTGCTTACAGAACTCTCTACGCCAATCGC 1143  
QY 1248 CTCCTCCGCTCATCCGTAGAGACACGAGACCCCTCATCGAGAGCTCAATCCCGGC 1307  
Db 1144 CTCCTCCGCTCATCCGTAGAGACACGAGACCCCTCATCGAGAGCTCAATCCCGGT 1203  
QY 1308 GACGCTTTGGAGCTTAGAGGCGGGGACACAGGGGTGTGGTACCGACTTTCGACGAGAC 1367  
Db 1204 GACGCTTTAGAGCTTAGAGGCGGGGTACAGGGGCGGTAGTACCGACTTTCGATGGTAT 1263  
QY 1368 GGGATGTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCGGCTGTCCTCTTC 1427  
Db 1264 GGGATGTGGATCTCATCTTGTTCATCGAGAGTCCATGGCTCAGCGCTGCTGTCTTC 1323  
QY 1428 CGGGCAATCAGGGCTTCAACAACTGGCTGCGAGTGGTCCACGACCGCGGTTTGG 1487  
Db 1324 CGGGCAATCAGGGCTTTCAGCAAACTGGTTCGTTGGTGTACCTCGCACCCGGTTCGG 1383  
QY 1488 GCTTTTGGCAGGGGAGCTAAGTCTGTCTTACACAAAGAGTGGGCGCCACCTGAGG 1547  
Db 1384 GCTTTTGGCAGGGGCGCAAGGTTGTCTTACACAAAGAGTGGGCGCGACCTACCG 1443  
QY 1548 ATCATCGAGGGGCTCAGGCTACCTGTGTGAGATGGAGCCGCTGGCACACTTTGGCCTG 1607  
Db 1444 ATCATGTAGGGGCTTCCGGCTACCTGTGTGAAATGGAGCTGTGGCACATTTTGGCCTG 1503  
QY 1608 GGGAGGATGAAGCCAGCAGTGTGAGGTGACGTGCGCCAGATGGCAAGATGGTGGCGCG 1667  
Db 1504 GAGCGGATGAAGCCAGCAGTGTGAGGTGAGTGTGAGTGGCCAGATGGCAGATGGTGGCGGA 1563  
QY 1668 AACCTGGCGCAGGGGAGATGAATCAGTGTGGAGATCCTCTACCCCGGATGAGGAC 1727  
Db 1564 ACTGTGGCGCAGGAGATGAATCAGTGTGGAGATCCTCTACCCCGGATGAGGAC 1623  
QY 1728 ACATTCAGGACCCAGCCCTCAGTGTGGAGTGGCCAGGATTTCTCCAGCAGCAATGGC 1787  
Db 1624 AAATTCAGAACACAGCCCTCAGTGTGGAGTGGCCAGGATTTCTCCAGCAGCAATGGC 1683  
QY 1788 CATTCATGAGCAACCAATGAATGATCCAGTTCCCATTCGTGTGCCCTCGACAAAGCCC 1847  
Db 1684 CATTCATGAGCAACCAATGAATGATCCAGTTCCCATTCGTGTGCCCTCGACAAAGCCC 1743  
QY 1848 GTATGTGTCAACCTTATGGAAGTACAGTGTGGAGTGGCCAGCAAGAGTGTGAGTGGGGC 1907  
Db 1744 GTATGTGTCAACCTTATGGAAGTACAGTGTGGAGTGGCCAGCAAGAGTGTGAGTGGGGC 1803  
QY 1908 TAGAGCCCAACAGGATGGCACAGCTCGGTGG 1941  
Db 1804 TATGAACCAATGAAGACGGCACAGCTGTGTGG 1837

## RESULT 4

BC042687 2147 bp mRNA linear HTC 02-JAN-2003  
LOCUS Mus musculus, clone IMAGE:3157049, mRNA.  
DEFINITION BC042687  
ACCESSION BC042687.1 GI:27469363  
VERSION  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM  
REFERENCE 1 (bases 1 to 2147)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (02-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapsb-f@mail.nih.gov](mailto:cgapsb-f@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 6 Row: p Column: 3  
 This clone has the following problem: no 5' EST match.

## FEATURES

Location/Qualifiers  
 1. 2147  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3157049"  
 /tissue\_type="Mammary tumor. C3(1)-Tag model. Infiltrating  
 ductal carcinoma. 5 month old virgin mouse."  
 /clone\_lib="NCI CGAP\_Mam6"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 66.1%; Score 1438; DB 11; Length 2147;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-266; Mismatches 0; Indels 0; Gaps 0;  
 Matches 1438; Conservative 0; CGCCTTCTCGGCGGTGGCCACGTAACCGGACA 564  
 1 AGGAGATCTACTTCTCTCAACACCAATAATGCCCTTCTCGGCGGTGGCCACGTAACCGGACA 60  
 565 AGTTGTTCAAGTTCGGCAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAACGTGG 624  
 61 AGTTGTTCAAGTTCGGCAATAACCGGTGGGAAGACATCTCTGAGCGATGAGGTCAACGTGG 120  
 625 CCGTGTGGTGGCCAGCCTCTTTCGCGGAGCCTCTGTGGCCTGTGGACAGAAAGGCT 684  
 121 CCGTGTGGTGGCCAGCCTCTTTCGCGGAGCCTCTGTGGCCTGTGGACAGAAAGGCT 180  
 685 CTGGACGCTACTCTATCTACATTCGCAATTAACGCTTACGCTTACGCTTACGCTTACGCTT 744  
 181 CTGGACGCTACTCTATCTACATTCGCAATTAACGCTTACGCTTACGCTTACGCTTACGCTT 240  
 745 TCATTGAATGACCTTGAGCCAGTGA CTTCTCCGCGGCGCATCTGGCGCTCAGAGATG 804  
 241 TCATTGAATGACCTTGAGCCAGTGA CTTCTCCGCGGCGCATCTGGCGCTCAGAGATG 300  
 805 TGGCTCTGAGGCTGGGCTGAGCAATAATACAGGGGCGGAGCGCTCAGCGTGGGCCCCA 864  
 301 TGGCTCTGAGGCTGGGCTGAGCAATAATACAGGGGCGGAGCGCTCAGCGTGGGCCCCA 360  
 865 TCCTCAGCAGCAGTGGCTCGGATATCTCTCGACCAATGAGATGGGCTTAACTTCTCTTT 924  
 361 TCCTCAGCAGCAGTGGCTCGGATATCTCTCGACCAATGAGATGGGCTTAACTTCTCTTT 420  
 925 TCACAAACCGGGCGCATGGCACTTTGTGAGCAGCTGCGGCGCATCTGTGTTGGACGACC 984  
 421 TCACAAACCGGGCGCATGGCACTTTGTGAGCAGCTGCGGCGCATCTGTGTTGGACGACC 480  
 985 CCACACAGCATGGCGAGGTGCGCCTGCTGACTTCAACCGTATGGCAAGTGGACA 1044  
 481 CCACACAGCATGGCGAGGTGCGCCTGCTGACTTCAACCGTATGGCAAGTGGACA 540  
 1045 TCGTCTATGSCAACTGGAATGGCCCCACCGCCTCTATCTGCAAAATGAGCACCCATGGGA 1104  
 541 TCGTCTATGSCAACTGGAATGGCCCCACCGCCTCTATCTGCAAAATGAGCACCCATGGGA 600

QY	1105	AGGTCCGCTTCCGGGACATCCGCTCA	1164
Db	601	AGGTCCGCTTCCGGGACATCCGCTCA	660
QY	1165	TCATCACCGCGGACTTTGACAATGA	1224
Db	661	TCATCACCGCGGACTTTGACAATGA	720
QY	1225	ACCGCAGCTCTCTCAGCCAAACCGG	1284
Db	721	ACCGCAGCTCTCTCAGCCAAACCGG	780
QY	1285	TCATCGAGGAGCTCAATCCCGGAGC	1344
Db	781	TCATCGAGGAGCTCAATCCCGGAGC	840
QY	1345	TGGTGACCGACTTCGACGGAGAGCG	1404
Db	841	TGGTGACCGACTTCGACGGAGAGCG	900
QY	1405	TGGCTCAGCCGCTGTCCGTCTTCCG	1464
Db	901	TGGCTCAGCCGCTGTCCGTCTTCCG	960
QY	1465	TGGTGCCACGCAACCCGGTTTGGGG	1524
Db	961	TGGTGCCACGCAACCCGGTTTGGGG	1020
QY	1525	AGAAGAGTGGGCGCCACCTTGAGG	1584
Db	1021	AGAAGAGTGGGCGCCACCTTGAGG	1080
QY	1585	AGCCCGTGGGACACTTTGGCCTGG	1644
Db	1081	AGCCCGTGGGACACTTTGGCCTGG	1140
QY	1645	CAGATGGCAAGATGTTGAGCCGGA	1704
Db	1141	CAGATGGCAAGATGTTGAGCCGGA	1200
QY	1705	TCCTCTACCCCGGGATGAGGACAC	1764
Db	1201	TCCTCTACCCCGGGATGAGGACAC	1260
QY	1765	GATTCTCCACGACGAGAAATGGCA	1824
Db	1261	GATTCTCCACGACGAGAAATGGCA	1320
QY	1825	TCGTGTCCCTCGAGCAAGCCCGTA	1884
Db	1321	TCGTGTCCCTCGAGCAAGCCCGTA	1380
QY	1885	CCAAACAGAGTGCAGTCGGGCTAC	1942
Db	1381	CCAAACAGAGTGCAGTCGGGCTAC	1438

RESULT 5  
 BX460454  
 LOCUS  
 DEFINITION BX460454 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 CSDF014YH11 5-PRIME, mRNA sequence.  
 ACCESSION BX460454  
 VERSION BX460454.1 GI:31037049  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1022)  
 AUTHORS Li W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)





```
Query Match      36.7%; Score 798; DB 12; Length 841;
Best Local Similarity 97.0%; Pred. No. 2.3e-143;
Matches 813; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 3 GGAGGCTCGAGGCGCAGCCCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGCGG 62
D 3 GGAGGCTCGAGGCGCAGCCCGGACCGGGGCTGGAGCAAGCAGCGCGCGCGCGCGG 63
QY 63 AGAGGCGGCGAGCGCGCGCGCTTCCACGCGCCCTAGGCGCGCGGCGGAGCGGAGG 122
D 64 AGAGGCGGCGAGCGCGCGCGCTTCCACGCGCCCTAGGCGCGCGGCGGAGCGGAGG 123
QY 123 ATGCTCTCCGAGCGCTGACCCCGGATGCTCAGAGATGTTACGTTCTCTGCTCTGG 182
D 124 ATGCTCTCCGAGCGCTGACCCCGGATGCTCAGAGATGTTACGTTCTCTGCTCTGG 183
QY 183 TTTCTGCGCCATCACTAGAGGGGTCCACGCGGCTGAACCCATGTTCACTGAGTCAACCAAC 242
D 184 TTTCTGCGCCATCACTAGAGGGGTCCACGCGGCTGAACCCATGTTCACTGAGTCAACCAAC 243
QY 243 TCAGTCTGCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGCAGTT 302
D 244 TCAGTCTGCTCTGACTATGACAGTAATCCACCCAGCTCAACTATGTTGGCAGTT 303
QY 303 ACTGATGTGACCATGATGGGCTTTGAGATCGTCTGCGGGGTACAATGACCCCAAC 362
D 304 ACTGATGTGACCATGATGGGCTTTGAGATCGTCTGCGGGGTACAATGACCCCAAC 363
QY 363 CTGGTTCTGAAGTATGACCCGGGCGGAGGCGGTGTAACATCGCGGTGATGAGCGC 422
D 364 CTGGTTCTGAAGTATGACCCGGGCGGAGGCGGTGTAACATCGCGGTGATGAGCGC 423
QY 423 AGCTCACTCTACTAGCGGCTGCGGACCGGAGGGGAAACCGATTGGGTGACAGCTGC 482
D 424 AGCTCACTCTACTAGCGGCTGCGGACCGGAGGGGAAACCGATTGGGTGACAGCTGC 483
QY 483 GACATCGAGGGGACCGCGGGAGGAGATCTACTTCTCAACCAATATGCTTCTCG 542
D 484 GACATCGAGGGGACCGCGGGAGGAGATCTACTTCTCAACCAATATGCTTCTCG 543
QY 543 GGGGTGCGCCACGTACACCGCAAGTTCTTCAAGTTCCGCAATACCGGTGGGAGACATC 602
D 544 GGGGTGCGCCACGTACACCGCAAGTTCTTCAAGTTCCGCAATACCGGTGGGAGACATC 603
QY 603 CTGAGCGATGAGGTCAACGTGGCGCGGTGGTGGCCAGCGCTTTTTCGCGACCGCTGTG 662
D 604 CTGAGCGATGAGGTCAACGTGGCGCGGTGGTGGCCAGCGCTTTTTCGCGACCGCTGTG 663
QY 663 GCCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATTCGCAATTACGCTAC 722
D 664 GCCTGTGTGACAGAAAGGCTCTGGACGCTACTCTATCTACATTCGCAATTACGCTAC 723
QY 723 GGTAAATGTGGCCCTGATGCGCTCATTTAAATGAGCCCTGAGGCGAGTACCTCTCCCGG 782
D 724 GTACATGTGGCCCTGATGCGCTCATTTAAATGAGCCCTGAGGCGAGTACCTCTACCGG 783
QY 783 GGCATTCTGGCGCTCAGAGATGTGGCTGTGAGCTGGGTGAGCAATATACAGGGG 840
D 784 GGCATTCTGGCGCTCAGAGATGTGGCTGTGAGCTGGGTGAGCAATATACAGGGG 841

RESULT 7
BX451456 904 bp mRNA linear EST 22-MAY-2003
LOCUS BX451456 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0DF014VH11 5-PRIME, mRNA sequence.
ACCESSION BX451456
VERSION BX451456.1 GI:31022247
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

## REFERENCE

1 (bases 1 to 904)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL

## COMMENT

Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5247.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAF012ZE12\_AF01150\_1&cluster=5247.r.  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BAF012ZE12\_AF01150\_1.

## FEATURES

Location/Qualifiers  
1..904  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF014VH11"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

## ORIGIN

Query Match 36.3%; Score 790.8; DB 13; Length 904;  
Best Local Similarity 96.8%; Pred. No. 5.9e-142;  
Matches 860; Conservative 0; Mismatches 22; Indels 6; Gaps 5;  
QY 838 GGGCGCGAGCGCTCAGCGTGGGCGCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGGC 897  
D 1 GGGCGCGAGCGCTCAGCGTGGGCGCCATCTCTCAGCAGCAGTGCCTCGGATATCTTCTGGC 59  
QY 898 ACAATGAGATGGCTTAACCTCTCTTCCACACCGGGCGATGGCACCTTTGTGACG 957  
D 60 ACAATGAGATGGCTTAACCTCTCTTTC-CAACCGGGCGATGGCACCTTTGTGACG 118  
QY 958 CTCGCGCGATGCTGTGTGGACGACCCCAACAGATGGCGGAGTGTGCGCTGCTG 1017  
D 119 CTCGCGCGATGCTGTGTGGACGACCCCAACAGATGGCGGAGTGTGCGCTGCTG 178  
QY 1018 ACTTCAACCGTGTGCAAAAGTGGACATCGTCTATGGCAACTGGATGGCGCCACCGCC 1077  
D 179 ACTTCAACCGTGTGCAAAAGTGGACATCGTCTATGGCAACTGGATGGCGCCACCGCC 238  
QY 1078 TCTATCTGCAAAATGAGCACCCATGGGAAGTTCGCTTCCGGGACATCGCTCACCCCAAGT 1137  
D 239 TCTATCTGCAAAATGAGCACCCATGGGAAGTTCGCTTCCGGGACATCGCTCACCCCAAGT 298  
QY 1138 TCTCCATGCCCCCTCTGTCGCGACGTCATCACC CGCGACTTTGACAATGACGAGGAGC 1197  
D 299 TCTCCATGCCCCCTCTGTCGCGACGTCATCACC CGCGACTTTGACAATGACGAGGAGC 358  
QY 1198 TGGAGATCTTCTTCAACAACATTGCTTACCGAGTCTCTCAGCGCAACCGCTCTTTCGCG 1257  
D 359 TGGAGATCTTCTTCAACAACATTGCTTACCGAGTCTCTCAGCGCAACCGCTCTTTCGCG 418  
QY 1258 TCATCCGTGAGAGCAGGAGACCCCTCATCAGGAGTCAATCCGCGGACGCGCTTGG 1317  
D 419 TCATCCGTGAGAGCAGGAGACCCCTCATCAGGAGTCAATCCGCGGACGCGCTTGG 478  
QY 1318 AGCCTGAGGCGCGGGGACAGGGGGTGTGGTACCGACTTCGACGAGAGCGGATGTGG 1377  
D 479 AGCCTGAGGCGCGGGGACAGGGGGTGTGGTACCGACTTCGACGAGAGCGGATGTGG 538  
QY 1378 ACCTCATCTTGTCCCATGGAGAGTCCATGCGCTCAGCGGCTGTCCGTCTTCCGGGGCAATC 1437



```

Db      539  ACCTCATCTTGTCCCATGAGAGTCCATGCTCAGCCGTGCTTCCTCCGGGCAATC 598
QY      1438  AGGGCTTCAACAACTAGCTGTGAGTGTGCGACGACACCCGGTTTGGGGCTTTGGCCA 1497
Db      599  AGGGCTTCAACAACTAGCTGTGAGTGTGCGACGACACCCGGTTTGGGGCTTTGGCCA 658
QY      1498  GGGGAGCTAAGTCTGTCTACACCAAG-AGAGTGGGGCCACCTGAGGATCATCGAC 1556
Db      659  GGGGAGCTAAGTCTGTCTACACCAAGAGAGTGGGGCCACCTGAGGATCATCGAC 718
QY      1557  GGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACATTTGGCCCTGGGGAGGAT 1616
Db      719  GGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACATTTGGCCCTGGGGAGGAT 778
QY      1617  GAAGCCAGAGTGTGAGGT--GAGTGGCCAGATGGCAAGATGTGAGCCGGAAGCTGG 1674
Db      779  GAAGCCAGAGTGTGAGGTGACCGTGGCCAGATTTGGCAGATGTGAGCCGGAAGCTGG 838
QY      1675  CCAGCGGGGAGATGAACCTCAGTGTGG-AGATCCCTACCCCGGGAT 1721
Db      839  CCAGCGGGGAGAGAACTAAGTGTGGAGAAATCTTTACCCCGGGAT 886

RESULT 8
BI550776      781 bp      mRNA      linear      EST 05-SEP-2001
LOCUS      603195670F01 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5275168 5',
DEFINITION      mRNA sequence.
ACCESSION      BI550776
VERSION      BI550776.1 GI:15438088
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 781)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1694 row: 1 column: 17
High quality sequence stop: 763.
Location/Qualifiers
1..781
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5275168"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC 95"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gtcgag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTNN-3', size-selected for average
insert size 2.5 kb and normalized to 10^5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this
is a NIH_MGC Library."

FEATURES
source
```

```

Query Match      55.5%; Score 773.8; DB 12; Length 781;
Best Local Similarity 99.7%; Pred. No. 1.1e-138;
Matches 775; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      3  GGAGGCTCGAGGCTCAGCCCGGACCCGGGGCTGGGAGCAAGAGCGGCGCGCGCGCGCG 62
Db      4  GGAGGCTCGAGGCTCAGCCCGGACCCGGGGCTGGGAGCAAGAGCGGCGCGCGCGCGCG 63
QY      63  AGAGGGGCGGACGAGGCGCGCTTCCACGCCCCCTAGGCGGGCGGCGCGGAGCGGAGG 122
Db      64  AGAGGGGCGGACGAGGCGCGCTTCCACGCCCCCTAGGCGGGCGGCGGAGCGGAGG 123
QY      123  ATGGGCTCCGAGGCGCTGACCCCGGCGATGTCCAGGATGTTACCGTTCTCTGCTCTCG 182
Db      124  ATGGGCTCCGAGGCGCTGACCCCGGCGATGTCCAGGATGTTACCGTTCTCTGCTCTCG 183
QY      183  TTTCTGCCCATCATCTGAGGGGTCCAGCGGGCTGAACCCCATGTTCTCATCTGAGTCA 242
Db      184  TTTCTGCCCATCATCTGAGGGGTCCAGCGGGCTGAACCCCATGTTCTCATCTGAGTCA 243
QY      243  TCAGTTCTGCTCTCTGACTATGACAGTAAATCCACCCAGCTCAACTATGTTGTCAGT 302
Db      244  TCAGTTCTGCTCTCTGACTATGACAGTAAATCCACCCAGCTCAACTATGTTGTCAGT 303
QY      303  ACTGATGTGACCATGATGGGACTTTGAGATCGTGTGGCGGGGTACAAATGGACCAAC 362
Db      304  ACTGATGTGACCATGATGGGACTTTGAGATCGTGTGGCGGGGTACAAATGGACCAAC 363
QY      363  CTGGTTCTGAAGTATGACCGGGCCAGAAAGCGGGCTGGTGAACATCGCGGTGATGAG 422
Db      364  CTGGTTCTGAAGTATGACCGGGCCAGAAAGCGGGCTGGTGAACATCGCGGTGATGAG 423
QY      423  AGCTCACCTCTACTACGCGCTCGGGACCGGCGAGGGGAACGCGCATTTGGGGTCA 482
Db      424  AGCTCACCTCTACTACGCGCTCGGGACCGGCGAGGGGAACGCGCATTTGGGGTCA 483
QY      483  GACATCGAGCGGGACCGGGCGGAGGAGATCTATCTTCTCAACACCAATATGCTTCT 542
Db      484  GACATCGAGCGGGACCGGGCGGAGGAGATCTATCTTCTCAACACCAATATGCTTCT 543
QY      543  GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGTGGGAGACAT 602
Db      544  GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGTGGGAGACAT 603
QY      603  CTGAGCGATGAGTCAACGCTGCCGTGTGTGGCCAGCGCTCTTTCCCGGAGCGCTCT 662
Db      604  CTGAGCGATGAGTCAACGCTGCCGTGTGTGGCCAGCGCTCTTTCCCGGAGCGCTCT 663
QY      663  GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTATCATTTGCCAATTAC 722
Db      664  GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTATCATTTGCCAATTAC 723
QY      723  GGTATATGGGGCTGTGATGCCCTCATTTGAATGGACCCCTGAGGCCAGTACCTCTCC 779
Db      724  GGTATATGGGGCTGTGATGCCCTCATTTGAATGGACCCCTGAGGCCAGTACCTCTCC 780

RESULT 9
BQ067249
LOCUS      BQ067249
DEFINITION      AGENCOURT 6767325 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5751633
5', mRNA sequence.
ACCESSION      BQ067249
VERSION      BQ067249.1 GI:19896295
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1101)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
REFERENCE
AUTHORS
TITLE
JOURNAL
```

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-i@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM12784 row: i column: 10  
High quality sequence stop: 612.

FEATURES  
source  
1. .1101  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5751633"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_115"  
/notes="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 35.5%; Score 773.6; DB 13; Length 1101;  
Best Local Similarity 93.9%; Pred. No. 1.3e-138;  
Matches 862; Conservative 0; Mismatches 41; Indels 15; Gaps 5;

QY 720 TACGGTAATGTGGCCCTGATGCCCTCATTTGAATGACCTGAGCCAGTGACCTCTCC 779  
Db 1 TACGGTAATGTGGCCCTGATGCCCTCATTTGAATGACCTGAGCCAGTGACCTCTCC 60  
QY 780 CGGGGCATCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGTGACGAATATACAGG 839  
Db 61 CGGGGCATCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGTGACGAATATACAGG 120  
QY 840 GCGCGAGCGTACGCTGGGCGCCCATCTCAGCAGCAGTGCTCGATATCTTCTGCGAC 899  
Db 121 GCGCGAGCGTACGCTGGGCGCCCATCTCAGCAGCAGTGCTCGATATCTTCTGCGAC 180  
QY 900 AATGAGATGGGCTAACTTCTTTTCCACACCGGGCGATGCGACCTTTGTGGACGT 959  
Db 181 AATGAGATGGGCTAACTTCTTTTCCACACCGGGCGATGCGACCTTTGTGGACGT 240  
QY 960 GCGCGCAGTCTGCTGTGGACGACCCACCAGCATGGCGAGGTGTCGCCCTGGCTGAC 1019  
Db 241 GCGCGCAGTCTGCTGTGGACGACCCACCAGCATGGCGAGGTGTCGCCCTGGCTGAC 300  
QY 1020 TTCAACCGTGTGCAAAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCCCAGCGCTC 1079  
Db 301 TTCAACCGTGTGCAAAAGTGGACATCGTCTATGGCAACTGGAAATGGCCCCCAGCGCTC 360  
QY 1080 TATCTGCAATGAGCACCCATGGAAGTCGGCTTCGGGACATCGCCTCAGCCCAAGTTC 1139  
Db 361 TATCTGCAATGAGCACCCATGGAAGTCGGCTTCGGGACATCGCCTCAGCCCAAGTTC 420  
QY 1140 TCCATGCCCTCCCTGTCCGACAGGTGATCACCGCCGACTTTGACAAATACCAAGAGCTG 1199  
Db 421 TCCATGCCCTCCCTGTCCGACAGGTGATCACCGCCGACTTTGACAAATACCAAGAGCTG 480  
QY 1200 GAGATCTTCTTAACACATTGCTACCGAGCTCTTACGCCAACCGCTCTTCCCGCTC 1259  
Db 481 GAGATCTTCTTACACACATTGCTACCGAGCTCTTACGCCAACCGCTCTTCCCGCTC 540  
QY 1260 ATCCGTAGAGCAGCGAGACCCCTCATCGAGAGCTCATCCCGGAGCGCTTGAG 1319

Db 541 ATCCGTAGAGCAGCAGGAGACCCCTCATCGAGAGCTCAATCCCGGAGCGCCTTGAG 600  
QY 1320 CCTGAGGCGCGGGGCACAGGGGGTGTGTGACCGACTTTCAGCGAGAGCGGATCTGGAC 1379  
Db 601 CCTGAGGCGCGGGGCACAGGGGGTGTGTGACCGACTTTCAGCGAGAGCGGATCTGGAC 660  
QY 1380 CTATCTTGTCCCATGGAGAGTCAATGGCTCAGCCGTGTCCTTCGCGGGCAATCAG 1439  
Db 661 CTATCTTGTCCCATGGAGAGTCAATGGCTCAGCCGTGTCCTTCGCGGGCAATCAG 720  
QY 1440 GGCTTCAACAACAACTGGCTGCGAGTGGTGCACGACCCCGGTTT---GGGGCCTTTGCC 1496  
Db 721 GGCTTCAACAACAACTGGCTGCGAGTGGTGCACGACCCCGGTTTGGGGCCTTTGCCA 780  
QY 1497 AGGGAGCTAAGTGTGCTTACACC-AAGAAGAGTGGGCGCCACC-----TGAGGATC 1550  
Db 781 GGGGAGCTAAAGTGTGCTTACACCACCAAGAAGTGGGCGCCACCCTGAGGGATCATC 840  
QY 1551 ATCAGCGGGGCTCAGGCTACCT---GTGTGAGATGGAGCCCGTGGCACACTTTGGCC-- 1605  
Db 841 GAACGGGGGCTCCAGCTTACCTGTGTGAGAAATGGAAGCCCGTGGCACACTTTTGGCCT 900  
QY 1606 -TGGGGAAGGATGAAGCC 1622  
Db 901 TGGGGAAGGATGAAGCC 918

RESULT 10  
BI758703  
LOCUS BI758703  
DEFINITION 603024078F1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5194530 5', mRNA sequence.  
ACCESSION BI758703  
VERSION BI758703.1 GI:15750294  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 770)  
NIH-MGC http://mgs.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-i@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11486 row: 1 column: 19  
High quality sequence start: 3  
High quality sequence stop: 770.

FEATURES  
source  
1. .770  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5194530"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_114"  
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH\_MGC Library."

ORIGIN



QY 544 GGGTGGCCACGTACACCGACCAAGTTGTTCAAGTTCGCCAAATAACCGGTGGGAAGACATCC 603  
Db |||||  
QY 542 GGGTGGCCACGTACACCGACCAAGTTGTTCAAGTTCGCCAAATAACCGGTGGGAAGACATCC 601  
Db |||||  
QY 604 TGAGCGATGAGTCAACCTGCGCGGTGTGGCCAGCCTCTTTGCCGAGCCT-CTGTG 662  
Db |||||  
QY 602 TGAGCGATGAGTCAACCTGCGCGGTGTGGCCAGCCTCTTTGCCGAGCCTGTGCTGTG 661  
Db |||||  
QY 663 GCCTGTG-TGGACAGAAAGGGCTC--TGGACCGTACTCTATC-TACATTGCCAATT-ACG 717  
Db |||||  
QY 662 GCCTGTGTTGGACCGAAAGGGCTCTCTGGACCGTACTCTATCTTACATTGCCAATTACG 721  
Db |||||  
QY 718 CCTAGGTAATGTGGCGCTGATGCCCTCATTTGAAATGGAACCTGAGCCGATGACCTCT 777  
Db |||||  
QY 722 CCTAGGTAATGTGGCGCTGATGCCCTCATTTGAATGG-GCCTGAGGCCAGTGACCTCT 780  
Db |||||  
QY 778 CCCGGGGCATCTGGCGCTC 797  
Db |||||  
QY 781 CCCGGGGCATCTGGCGCTC 800  
Db |||||

RESULT 12  
BI819378  
LOCUS 860 bp mRNA linear EST 04-OCT-2001  
DEFINITION 60304549F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5175469 5',  
mRNA sequence.  
ACCESSION BI819378  
VERSION BI819378.1 GI:15930928  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 860)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM11437 row: b column: 14  
High quality sequence stop: 824.  
FEATURES  
source  
Location/Qualifiers  
1..860  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5175469"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_115"  
/note="Organ: pooled brain, lung, testis; Vector:  
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb,  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH\_MGC library."

## FEATURES

## source

## Location/Qualifiers

## 1..860

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="IMAGE:5175469"

## /lab\_host="DH10B"

## /clone\_lib="NIH\_MGC\_115"

## /note="Organ: pooled brain, lung, testis; Vector:

## pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA

## source anonymous pool of 6 male brains, age range 23-27; 1

## male lung, age 27; and 1 male testis, age 69. Library is

## oligo-dT primed and directionally cloned (EcoRV site is

## destroyed upon cloning). Average insert size 1.8 kb,

## insert size range 1-3 kb. Library is normalized and

## enriched for full-length clones and was constructed by C.

## Gruber (Invitrogen). Research Genetics tracking code

## 021. Note: this is a NIH\_MGC library."

QY 1 CGGAGGGCTCAGGCCAGCCCGGACCCGGGGCTGGAGCAAGCAGCGCGCGCGCGCGG 60  
Db |||||  
QY 13 CGGAGGGCTCAGGCCAGCCCGGACCCGGGGCTGGAGCAAGCAGCGCGCGCGCGG 72  
Db |||||  
QY 61 GCAGAGGGCGGAGGAGCGCC-CGCTTCCACGCCCTAGCGGGCGGGCGGAGAGCGGG 119  
Db |||||  
QY 73 GCAGAGGGCGGAGGAGCGCGCTTCCACGCCCTAGCGGGCGGGCGGAGAGCGGG 132  
Db |||||  
QY 120 AGGATGGCTCCGAGCGCTGACCCCGGCGATGTCCAGGATGTTTACCGTTCTCTGTGCTG 179  
Db |||||  
QY 133 AGGATGGCTCCGAGCGCTGACCCCGGCGATGTCCAGGATGTTTACCGTTCTCTGTGCTG 192  
Db |||||  
QY 180 -TGGTTTCTGCCCATCTACTGAGGGGTCCAGGGGCTGAACCCATGTTTCACTGCACTG 238  
Db |||||  
QY 193 ATGGTTTCTGCCCATCTACTGAGGGGTCCAGGGGCTGAACCCATGTTTCACTGCACTG 252  
Db |||||  
QY 239 CAATCTCAGTTCTGCTCTGCTGATGACAGTAATCCACCCAGCTCAACTATGGTGTGGC 298  
Db |||||  
QY 253 CAATCTCAGTTCTGCTCTGCTGATGACAGTAATCCACCCAGCTCAACTATGGTGTGGC 312  
Db |||||  
QY 299 AGTTACTGATGTGGACCATGATGGGACTTTTGGATCGTGGCGGGGTACAATGGACC 358  
Db |||||  
QY 313 AGTTACTGATGTGGACCATGATGGGACTTTTGGATCGTGGCGGGGTACAATGGACC 372  
Db |||||  
QY 359 CAATCTGTTTCTGAGTAATGACCGGGCCAGAGCGGCTGGTGAACATCGCGGTGATGA 418  
Db |||||  
QY 373 CAATCTGTTTCTGAGTAATGACCGGGCCAGAGCGGCTGGTGAACATCGCGGTGATGA 432  
Db |||||  
QY 419 GCGCAGCTCACCCCTACTACGCGCTGCGGGACCGGAGGGGAACGCATTGGGTGTCAGC 478  
Db |||||  
QY 433 GCGCAGCTCACCCCTACTACGCGCTGCGGGACCGGAGGGGAACGCATTGGGTGTCAGC 492  
Db |||||  
QY 479 CTGGACATTCAGCGGGACGCGCGGAGGAGATCTACTTCTCAACACCAATAATGCTT 538  
Db |||||  
QY 493 CTGGACATTCAGCGGGACGCGCGGAGGAGATCTACTTCTCAACACCAATAATGCTT 552  
Db |||||  
QY 539 CTCGGGGTGGCCACGTTACACCGACAGTGTTCAGTTCGCGAATAACCGGTGGGAAGA 598  
Db |||||  
QY 553 CTCGGGGTGGCCACGTTACACCGACAGTGTTCAGTTCGCGAATAACCGGTGGGAAGA 611  
Db |||||  
QY 599 CATCTCAGCGCATGA-GGTCAACGTGGCCCTGGTG-TGGCCAGCTCTTTGCCGAGCGC 656  
Db |||||  
QY 612 CATCTCAGCGCATGAGGTCAACGTGGCCCGGGTGTGGCCAGCTCTTTGCCGAGCGC 671  
Db |||||  
QY 657 TCT-GTGGCTGTGTGG-ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTTGCCAATT 714  
Db |||||  
QY 672 TCTGTGGCTGTGTGGAAAGAGGGTCTTGGACGCTACTCTATCTACATTTGCCAATT 731  
Db |||||  
QY 715 ACG-CTTACGGTAAT-TTGGGCCCTGATGCCCTCATTTGAATGGACCTGA-GGCCAGTG 771  
Db |||||  
QY 732 ACGCCCTACGGTAATTTGGTGGCCCTGATGCCCTCATTTGAATGGACCTGAGGGCCAGTG 791  
Db |||||  
QY 772 ACCTCTCCCGGGCATTTCTGGCGCTCAGAGATGGCTGTGAGGCTGGGTGAGCAAT 831  
Db |||||  
QY 792 ACCTCTCCCGGGCATTTCTGGCGCTCAGAGATGGCTGTGAGGCTGGGTGAGCAAT 851  
Db |||||  
QY 832 ATACAGGGG 840  
Db |||||  
QY 852 ATACGGGCG 860  
Db |||||

## RESULT 13

## BF527664

## LOCUS

## 602040644F1 NCI\_CGAP\_Brn67 Homo sapiens cDNA clone IMAGE:4178363

## DEFINITION

## 5', mRNA sequence.

## ACCESSION

## BF527664

## VERSION

## BF527664.1

## KEYWORDS

## EST.

## SOURCE

## Homo sapiens (human)

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 31.8%; Score 692.2; DB 12; Length 860;  
Best Local Similarity 96.1%; Pred No. 6e-123;  
Matches 816; Conservative 0; Mismatches 23; Indels 10; Gaps 10;

## ORIGIN

## Query Match

## Best Local Similarity

## Matches 816; Conservative

## 0; Mismatches 23; Indels 10; Gaps 10;



Db 4 GGAGGCTCGAGGCGAGCCCGGAGCCGCGGCTGGAGCAGCAGCGCGCGCGCGCGC 63  
QY 63 AGAGCGGCGAGCGAGCGCGCTTCCACGCCCTAGCGCGCGCGCGCGAGCGGAGG 122  
Db 64 AGAGCGGCGAGCGAGCGCGCTTCCACGCCCTAGCGCGCGCGCGCGAGCGGAGG 123  
QY 123 ATGGCTCCGAGCGCTGACCCCGGCGATGTCAGAGATGTTACCGTTCCTGCTGCTCTCG 182  
Db 124 ATGGCTCCGAGCGCTGACCCCGGCGATGTCAGAGATGTTACCGTTCCTGCTGCTCTCG 183  
QY 183 TTTCTGCCCATCACTAGAGGGTCCAGCGGGTGAACCCATGTTCACTGAGTCAACCAAC 242  
Db 184 TTTCTGCCCATCACTAGAGGGTCCAGCGGGTGAACCCATGTTCACTGAGTCAACCAAC 243  
QY 243 TCAGTTCGCTCTGACTATGACAGTAATCCACCCAGCTCACTATGTTGGCAGTT 302  
Db 244 TCAGTTCGCTCTGACTATGACAGTAATCCACCCAGCTCACTATGTTGGCAGTT 303  
QY 303 ACTGATGTGACCATGATGGGGAATTGAGATCGTCTGGCGGGTACAATGGACCCCAAC 362  
Db 304 ACTGATGTGACCATGATGGGGAATTGAGATCGTCTGGCGGGTACAATGGACCCCAAC 363  
QY 363 CTGCTTCTGAAGTATGACCGGGCCAGAGCGGTGTGAACATCGCGGTGCGATGAGCGC 422  
Db 364 CTGCTTCTGAAGTATGACCGGGCCAGAGCGGTGTGAACATCGCGGTGCGATGAGCGC 423  
QY 423 AGCTCACTTACTACGCGCTCGGGACCGGAGGAGCGCATGTTGGGTACAGACCTGCG 482  
Db 424 AGCTCACTTACTACGCGCTCGGGACCGGAGGAGCGCATGTTGGGTACAGACCTGCG 483  
QY 483 GACATCGACGGGAGCGCGGAGAGATCTACTTCTTCAACCAATATGCTTCTCG 542  
Db 484 GACATCGACGGGAGCGCGGAGAGATCTACTTCTTCAACCAATATGCTTCTCG 543  
QY 543 GGGGTGCGACGATACCGCAAGTGTTCGCAATATACCGGTGGGAAGATC 602  
Db 544 GGGGTGCGACGATACCGCAAG-TGTTCAAGTTCGCAATATACCGGTGGGAAGATC 602  
QY 603 CTGAGCGATGAGTCAA-CGTGGCCGCTGGGTGCGCAGCGCTTTG-CGGAGCGCTCG 660  
Db 603 CTGAGCGATGAGTCAA-CGTGGCCGCTGGGTGCGCAGCGCTTTG-CGGAGCGCTCG 662  
QY 661 TGGCTGTGTGGACAGAAAGGCTCTGG-AGCTACTCTA---TCTAATGSCCAATPAC 716  
Db 663 TGGCTGTGTGGACAGAAAGGCTCTGGAAAGCTCTTACTTCTAATTTGCCAATPAC 722  
QY 717 GCCTACGTAATGT-----GGGCCCTGATGCCCT-CATTGAATGACCCCTGAGGCCAGTG 771  
Db 723 GCCCTACGTAATGTGTGGCGCCCTGATGCCCTCCGTTGCAATGACCCCTGAGGGCAGTT 782  
QY 772 AC---CTCTCCCGGGGCACTTGGCGCTC--AGAGATGTGCTGTCTGA--GGCTGGGCTC 824  
Db 783 GCCCTTACCGGGGCACTTGGCGCTCAGAGGATTTGTTGTTGAGCGCTGGTCA 842  
QY 825 AGCAAAATATACAGGGGCGGAGCGCTCAGCGTGGGCCCC-----ATCCTCAGCAGCAG-T 878  
Db 843 GGCATAATTACCGGGGCGGAGCGCTACACAGGGGCCCCGATGCTCTCAGCAGAGAT 902  
QY 879 GCCTCGGATATCTTCTGCGAATGAGAAATGGGCGCTAACTCTCTTTTCCACACCGGGGC 938  
Db 903 GCGCTCGGAATTCGCTGCGCAAGTACAAAGGGGCGTAATCCTCTGCTCCACACAGGGCGGAGC 962  
QY 939 G 939  
Db 963 G 963

RESULT 15  
BQ183029/c  
LOCUS  
DEFINITION  
BQ183029  
UI-H-EUO-azs-e-03-0-UI.s1 NCI CGAP\_Carl Homo sapiens cDNA clone  
IMAGE:5852906 3', mRNA sequence.

ACCESSION BQ183029  
VERSION BQ183029.1  
KEYWORDS GI:20358579  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 656)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/nciogap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-femail.nih.gov  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Seq primer: M13 FORWARD  
POLYA=Yes.  
FEATURES  
Location/Qualifiers  
1..656  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5852906"  
/tissue\_type="Osteoarthritic Cartilage"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP Carl"  
/notes="Organ: Knee; Vector: pRT3-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
NCI CGAP Carl is a cDNA library containing the following  
tissue(s): Osteoarthritic Cartilage The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pRT3-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
TGATCAGCT.  
TAG\_TISSUE=osteoarthritic cartilage  
TAG\_LIB=UI-H-EUO  
TAG\_SEQ=TGATCAGCT"  
ORIGIN  
Query Match 29.5%; Score 642.4; DB 13; Length 656;  
Best Local Similarity 99.8%; Pred. No. 2.2e-113;  
Matches 643; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1534 GGGCCCACTGAGGATCATCGCGGGCTCAGGCTACTGTGTGAGATGGAGCCCGTGG 1593  
Db 651 GGGCCCACTGAGGATCATCGCGGGCTCAGGCTACTGTGTGAGATGGAGCCCGTGG 592  
QY 1594 CACACTTTGGCTCGGAGGATGAAGCCAGTGTGAGGTGACGTGGCCAGATGCA 1653  
Db 591 CACACTTTGGCTCGGAGGATGAAGCCAGTGTGAGGTGACGTGGCCAGATGCA 532  
QY 1654 AGATGTTGAGCCGGAACGTGGCCAGCGGAGATGAACCTCAGTCTGAGATCTCTTACC 1713  
Db 531 AGATGTTGAGCCGGAACGTGGCCAGCGGAGATGAACCTCAGTCTGAGATCTCTTACC 472  
QY 1714 CCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAGATTTCTCC 1773  
Db 471 CCGGGATGAGGACACACTTCAGGACCCAGCCCACTGGAGTGTGGCCAGATTTCTCC 412  
QY 1774 AGCAGAAATGGCCATTCATGGACCAATGATGCATCCAGTTCCTCCATTCGTGTGCC 1833

```

Db      411 AGCAGGAAAATGGCCATTGCGATGACACACCAATGAATGCATCCAGTTCCCATTCGTGTGCC 352
QY      1834 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGTGCCTGACCAACAAGA 1893
Db      351 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGTGCCTGACCAACAAGA 292
QY      1894 AGTGCAGTTCGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCC 1953
Db      291 AGTGCAGTTCGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCC 232
QY      1954 CTGTGTTGAAGATAGTGAACACCAAGTTGGGAAGAGCCTTGTCCCTGAATCACTGAAT 2013
Db      231 CTGTGTTGAAGATAGTGAACACCAAGTTGGGAAGAGCCTTGTCCCTGAATCACTGAAT 172
QY      2014 CACTGCTTTGAATCACCCTGCTGATACCTGTTGATCAGGAACACTTACCTGGAACCTCA 2073
Db      171 CACTGCTTTGAATCACCCTGCTGATACCTGTTGATCAGGAACACTTACCTGGAACCTCA 112
QY      2074 CTGAGCAGGATACAAACTTCTATTGTATTAAAGCTATTATACATTAAAGATTGGGGGTGC 2133
Db      111 CTGAGCAGGATACAAACTTCTATTGTATTAAAGCTATTATACATTAAAGATTGGGGGTGC 52
QY      2134 TACCTTACATATAAATTCCCATTTCTCTTGAAAAA 2177
Db      51 TACCTTACATATAAATTCCCATTTCTCTTGAAAAA 8

```

Search completed: August 9, 2004, 21:15:51  
 Job time : 5112 secs